PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) Internati nal Patent Classification 6 :
C12Q 1/68, C12P 19/34, G01N 21/00

A1

(11) International Publication Number: **WO 95/21269**

(43) International Publication Date: 10 August 1995 (10.08.95)

(21) International Application Number: PCT/US95/01395

(22) International Filing Date: 2 February 1995 (02.02.95)

(30) Priority Data:
08/192,491 4 February 1994 (04.02.94) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/192,491 (CIP)
Filed on 4 February 1994 (04.02.94)

(71)(72) Applicants and Inventors: PERLIN, Mark, K. [US/US]; 5904 Beacon Street, Pittsburgh, PA 15217 (US). GORIN, Michael, B. [US/US]; 110 Crofton Drive, Pittsburgh, PA 15238 (US).

(74) Agent: SCHWARTZ, Ansel, M.; Suite 301, 425 N. Craig Street, Pittsburgh, PA 15213 (US).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: METHOD AND APPARATUS FOR ANALYZING GENETIC MATERIAL

(57) Abstract

Modern medicine exploits relatively little of an individual's genetic composition in directing preventative, diagnostic, or therapeutic interventions. However, tracing the descent of chromosomal segments within families and populations, and then correlating with phenotypic traits, would enable accurate assessment of risk for common multifactorial diseases. This information could then be used to customize medical interventions to the major medical conditions for which an individual had significant risk. The primary obstacle has been the very many genotyping experiments and computations required to densely sample genomes. The invention pertains to a system that enables high-throughput genotyping, and thus the effective determination of such risks, and other useful genetic information. The inventi n also pertains to methods for determining the size of simple tandem repeat alleles by nucleic acid hybridization, including forming mismatched heteroduplexes and quantitating their single-stranded loop sizes.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

METHOD AND APPARATUS FOR ANALYZING GENETIC MATERIAL

FIELD OF THE INVENTION

The present invention pertains to a process for determining inheritance patterns in eukaryotic DNA. More specifically, the present invention is related to densely sampling the genome with polymorphic genetic markers using a hybridization-based genotyping method, and then using this genetic information to assess the trait inheritance, including disease susceptibility, mendelian genetic disorders, and complex traits relevant for plant or animal husbandry. One such hybridization-based genotyping method entails forming mismatched heteroduplexes and quantitating single-stranded loop sizes.

BACKGROUND OF THE INVENTION

The specific objective of the system is genome-wide high-resolution genotyping for the purpose of health risk assessment, including genetic susceptibility for disease, and identification of disease-associated genes. The means for achieving this is genotyping polymorphic genetic loci by hybridization assays.

In meiotic recombination, large regions of parental chromosomes are interleaved and passed on to the next generation. By effecting a very dense sampling of the genome (i.e., all the chromosomes) for every individual in a large family, one can determine who has inherited which portions of which chromosomes from whom. That is, the dense sampling serves to tag the origin and descent of linear chromosomal fragments throughout the pedigree. By correlating the

genotypic inheritance pattern of chromosomal fragments with the phenotypic occurrence of common multifactorial disease in individuals, culprit chromosomal regions can be identified. From this analysis, accurate risk assessments can be made for individuals based on their genotype, in the context of their entire kinship. Genome mismatch scanning (Nelson, S.F., McCusker, J.H., Sander, M.A., Kee, Y., Modrich, P., and Brown, P.O. 1993. Genomic mismatch scanning: a new approach to genetic linkage mapping. *Nature Genetics*, 4(May): 11-18.), incorporated by reference, is one such approach, but has limited throughput since experiments are done on pairs (not sets) of individuals.

For 1 centiMorgan (cM) resolution genome sampling, about three thousand highly polymorphic genetic loci would be required for a medium-resolution genome-wide genotyping. High resolution at 0.1cM would therefore require genotyping no more than 30,000 genetic loci. Currently, as part of the world-wide Human Genome Project (Watson, J.D., Gilman, M., Witkowski, J., and Zoller, M. 1992. *Recombinant DNA, Second Edition*. New York, New York: W.H. Freeman and Company), incorporated by reference, roughly 30,000 highly polymorphic genetic sequence tagged site (STS) (Olson, M., Hood, L., Cantor, C., and Botstein, D. 1989. A common language for physical mapping of the human genome. *Science*, 245: 1434-35.), incorporated by reference, loci will be developed and mapped in the next three years. A sequence-tagged site is defined herein as a location on a genome characterized by at least one sequence. Much of this effort is done by Weissenbach's group at CEPH in France (Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., and Lathrop, M. 1992. A second generation linkage map of the human genome. *Nature*, 359: 794-801),

incorporated by reference, and by Lander's group at the Whitehead Institute in Cambridge, Massachusetts. STSs ar readily amplified by means of the polymerase chain reaction (PCR). These STSs will largely take the form of variable nucleotide tandem repeat (VNTR) sequences (Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, I., and White, R. 1987. Variable number tandem repeat (VNTR) markers for human gene mapping. *Science*, 235: 1616-1622.), incorporated by reference,that have several nucleotides repeated a fixed (though highly polymorphic) number of times at any allele.

Importantly, the approach described herein centers on a detailed examination of such highly polymorphic intron genetic markers, rather than the highly conserved genes and their exon coding regions. However, the method also applies to expanded repeats within genes, and specific nucleotide alterations of specific DNA sequences.

Achieving this goal requires genome-wide high-resolution genotyping (1) an associated technology that will reduce the cost and error of the requisite genotyping, and thus enable widespread usage. Further, this technology must be coupled with (2) data acquisition and analysis methods that allow for fully automated error detection, risk analysis, and linkage analysis for both populations and families. Completion of this analysis generates a vast amount of data, hence the results must (3) be presented in a targeted fashion to disparate groups of end-users.

Much of the following description focuses on task (1), the novel parallel genotyping apparatus for polymorphic VNTRs. The approach is to spatially localize each genetic

locus in a two-dimensional array, and then locally aggregat PCR-amplified DNA products to the proper array regions. Then, perform DNA hybridization studies by means of a detection mechanism to quantitate properties of the PCR products, and thereby determine the alleles (i.e., the genotype) for every genetic locus.

More precisely, a VNTR is a linear sequence of (deoxy)nucleotides of the pattern LW_nR, where W is a short DNA sentence repeated n times, contained within two flanking regions of unique sequences: the left flanking region L, and the right flanking region R. These flanking sequences establish the singularity of a specific VNTR within a haploid genome. These unique sequences allow a VNTR to be associated with a specific location within the genome such that it can be physically or genetically mapped with respect to other DNA markers and/or genetic traits and disorders. Variations in the number of repetitive elements within the VNTR are common among individuals and allow specific alleles to be tracked as they are genetically transmitted from individuals to their offspring.

An important subclass of VNTRs is the short tandem repeat (STR), where n tends to be small (e.g., < 100), and repeating unit short (e.g., between two and five). For example, a CA-repeat is an STR where the dinucleotide CA is repeated n times, where n ranges in a human population from roughly ten to forty. There are an estimated 100,000 such CA-repeat loci in the human genome. Other VNTRs include trinucleotide and tetranucleotide repeats. Following PCR, the allelic variation in tandem repeat number can be

determined by DNA size measurements using polyacrylamide gel electrophoresis.

These STRs and VNTRs are important for several reasons. (1) Many VNTRs have been associated with specific diseases (e.g., Huntington's disease, fragile X syndrome) (Kremer, I., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R., and Richards, R.I. 1991. Mapping of DNA instability at the Fragile X to a trinucleotide repeat sequence $p(CCG)_n$. *Science*, 252: 1711-1714), incorporated by reference, where, in "anticipation", larger n often correlates with increased severity. (2) STRs serve as highly useful markers for specific diseases (Clemens, P., Fenwick, R., Chamberlain, J., Gibbs, R., de Andrade, M., Chakraborty, R., and Caskey, C. 1991. Linkage analysis for Duchenne and Becker muscular dystrophies using dinucleotide repeat polymorphisms. *Am J Hum Genet*, 49: 951-960.), incorporated by reference. (3) STRs are useful as sequence tagged sites (STSs) (Olson, M., Hood, L., Cantor, C., and Botstein, D. 1989. A common language for physical mapping of the human genome. *Science*, 245: 1434-35.), incorporated by reference, in physical mapping studies. (4) There is tremendous genetic polymorphism at these loci (Weber, J., and May, P. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet*, 44: 388-396.), incorporated by reference. For each polymorphic locus, n may assume a wide range of allelic values in the population. Therefore, STRs are highly polymorphic loci that can be used in genetic linkage (Ott, J. 1991. *Analysis of Human Genetic Linkage, Revised Edition*. Baltimore, Maryland: The Johns Hopkins University Press.), incorporated by reference, and chrom some

fingerprinting (Jeffreys, A.J., Wilson, V., and Thein, S.L. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature*, 314: 67-73. Jeffreys, A.J., Wilson, V., and Thein, S.C. 1985. Individual-specific fingerprints of human DNA. *Nature*, 316: 76-78.), incorporated by reference, studies that densely sample the genome.

Since STRs are easily amplified via PCR (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. 1990. *PCR Protocols: A Guide to Methods and Applications*. San Diego, CA: Academic Press. Mullis, K.B., Faloona, F.A., Scharf, S.J., Saiki, R.K., Horn, G.T., and Erlich, H.A. 1986. Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.*, 51: 263-273. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, B.T., Mullis, K.B., and Erlich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239: 487-491.), incorporated by reference, and (by definition) their alleles differ only in the repeat number n, genotyping is easily effected by measuring the total length of the PCR product. This is commonly done by spatially (or temporally) separating DNA molecules of different sizes (or conformations) using, for example, gel electrophoresis. Other well-known approaches include mass spectroscopy, denaturing gradient gel electrophoresis, and chemical assays. A newer gel-based approach is two-dimensional DNA typing (te Meerman, G.J., Mullaart, E., van der Meulen, M.A., den Daas, J.H.G., Morolli, B., Uitterlinden, A.G., and Vijg, J. 1993. Linkage analysis by two-dimensional DNA typing. *Am. J. Hum. Genet.*, 53: 1289-1297.), incorporated by reference. However, these measurements all have associated costs. In particular, none

are particularly cost effective for genotyping the thousands of STR loci that are needed for densely sampling genomes.

This invention therefore describes more cost effective approaches that enable higher throughput STR genotyping. These methods employ nucleotide hybridization assays that directly measure the number of STR repeat units, rather than total fragment length. Such detections by hybridization are miniaturizable, hence parallelizable (Monaco, A.P., Lam, V.M.S., Zehetner, G., Lennon, G.G., Douglas, C., Nizetic, D., Goodfellow, P.N., and Lehrach, H. 1991. Mapping irradiation hybrids to cosmid and yeast artificial chromosome libraries by direct hybridization of Alu-PCR products. *Nucleic Acids Res*, 19(12): 3315-3318.), incorporated by reference, and, ultimately, highly manufacturable. Further, they can be adapted to work in chemical solutions, or on substrates with small surface area.

Two novel methods for STR allele determination at a locus are introduced, both based on genotyping by hybridization. The first method entails creating and detecting loop mismatches in heteroduplexes formed from the alleles' PCR products. The second method uses hybridization panels to determine the alleles.

SUMMARY OF THE INVENTION

The present invention pertains to an apparatus for analyzing the genetic material of an organism. The apparatus comprises means for amplifying the genetic material of the organism. The apparatus also comprises means for characterizing the amplified genetic material. The characterizing means is in communication with the amplifying

means. The characterizing means contains all of the genetic material within a region having a radius of less than two feet. The amplifying means and characterizing means characterize the genetic material at a rate exceeding 100 sequence-tagged sites per hour per organism. The sequence-tagged sites are inherent to the genetic material.

Preferably, the genetic material includes nucleotide sequences. The amplifying means preferably includes a reaction plate with which the genetic material is in contact. The reaction plate has a plurality of chambers, each of which is disposed in a unique location of the plate corresponding to a location within a genome having at least one nucleotide sequence. The characterizing means preferably includes means for detecting whether a chamber contains a nucleotide sequence of the genetic material corresponding to the chamber's unique location.

The apparatus preferably also includes a thermocycler in thermal communication with the plate to heat and cool the plate. The detecting means preferably includes a detector connected to the chambers which produces a chamber signal for each chamber corresponding to genetic material in each chamber. The detecting means preferably also includes a processor in communication with the detector which receives the signal and identifies unique properties of the nucleotides in each chamber. The unique properties of the nucleotide of the genetic material in each chamber pertain to a number of nucleotides in any of the nucleotide sequences of the genetic material.

The amplifying means preferably includes at least one nucleotide sequence that corresponds to each chamber and

which is in contact with the chamber. Each nucleotide sequence interacts with the nucleotide sequence of the genetic material of the nucleotide sequence if it is present.

The present invention also pertains to a method for analyzing genetic material of an organism. The method comprises the steps of amplifying the genetic material. Then there is the step of characterizing the amplified genetic material in a region having a radius of less than 20 feet at a rate exceeding 100 sequence-tagged sites per hour per organism. Preferably, the genetic material includes RNA or DNA. After the characterizing step, there preferably is the step of accessing risk of illness for which there is a genetic susceptibility in the organism. Such illnesses can include cancer, heart disease, etc.

The present invention also pertains to a method for manufacturing an apparatus for analyzing genetic material of an organism. The method comprises the steps of placing corresponding sequence-tagged sites in contact with corresponding chambers of a plate. Then, there is the step of connecting detectors to the chambers which can detect where the nucleotide sequences of the genetic material of the organism, when placed in contact with the chambers, have reacted with the corresponding sequence-tagged sites in the corresponding chamber. Then, there is the step of placing a thermocycling device in contact with the plate to cause the sequence-tagged sites in the chambers to react with genetic material of the organism that is placed in contact with the chambers. Next, there is the step of connecting a computer to the detectors and to the thermocycling device to control operation of the thermocycling device, and to receive signals

which correspond to the genetic material of the organism and the sequence-tagged sites of each chamber from the detectors.

The present invention also pertains to a method for determining the size of nucleotide sequences of an STR marker contained on genetic material comprising the steps of: amplifying the nucleotide sequences of the genetic material in a region relating to the STR marker. Then there is the step of performing nucleic acid hybridizations on th amplified nucleotide sequences. Then there is the step of producing signals corresponding to the hybridizations of the amplified nucleotide sequences. Then there is the step of determining the sizes of the nucleotide sequences contained in the genetic material.

BRIEF DESCRIPTION OF THE DRAWINGS

In the accompanying drawings, the preferred embodiment of the invention and preferred methods of practicing the invention are illustrated in which:

Figure 1 is a schematic representation of a preferred embodiment of the apparatus.

Figure 2 is a schematic representation of parts of DNA molecules for name convention purposes.

Figures 3a-3d list the steps for parallel genotyping of the present invention.

Figures 4a and 4b are schematic representations of mismatched loops formed from allele DNA.

Figure 5 includes figures 5a-5c and is a schematic representation of loop mismatch for determining a sum of STR alleles.

Figure 6 includes figure 6a and is a block diagram showing loop mismatch for determining a difference of STR alleles.

Figure 7 is a flow chart for determining the STR alleles from the sum and difference.

Figure 8 is a flow chart of loop mismatch protocol for a single STR locus.

Figure 9 is a flow chart for reducing the number of PCR experiments.

Figures 10a-10c show representations for increasing measured signal from loops with respect to summation experiment.

Figures 11a and 11b are representations for increasing measured signal from loops with respect to difference experiments.

Figure 12 is a flow chart of concordance mapping for genetic patterns.

Figure 13 includes parts a-c and is a flow chart for determining an STR allele sum from a nucleic acid synthesis step.

Figure 14 includes parts a-c and is a flow chart for determining an STR allele difference from a nucleic acid synthesis step.

Figure 15 is a flow chart for determining STR alleles from a nucleic acid synthesis step.

Figure 16 is a schematic representation of an assay for determining STR alleles from a nucleic acid ligation step.

Figure 17 includes parts a-b and is a schematic representation of an assay for determining STR alleles from a nucleic acid loop ligation step.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawings wherein like reference numerals refer to similar or identical parts throughout the several views, and more specifically to figure 1 thereof, there is shown an apparatus for analyzing the genetic material of an organism. The apparatus comprises means for amplifying the genetic material of the organism. The apparatus also comprises means for characterizing the amplified genetic material. The characterizing means is in communication with the amplifying means. The characterizing means contains all of the genetic material within a region having a radius of less than two feet. It should be noted that the region could have a radius of any reasonable size commensurate with the requirements of the task. For instance, the radius of the region could range from 1 cubic millimeter up to 10 feet by and anywhere in between. The amplifying means and characterizing means characterize the

genetic material at a rate preferably exceeding 100 sequence-tagged sites per hour per organism. It should be noted that the rate could be up to 100,000 sequence-tagged sites per hour per organism, or as slow as desired, or any rate in between. Also, per organism could also be defined to be the characterization of genetic material of multiple organisms. The sequence-tagged sites are inherent to the genetic material.

Preferably, the genetic material includes nucleotide sequences. The amplifying means preferably includes a reaction plate 102 with which the genetic material is in contact. The reaction plate 102 has a plurality of chambers, each of which is disposed in a unique location of the plate 102 corresponding to a location within a genome having at least one nucleotide sequence. The characterizing means preferably includes means for detecting whether a chamber contains a nucleotide sequence of the genetic material corresponding to the chamber's unique location.

The apparatus preferably also includes a thermocycler 104 in thermal communication with the plate 102 to heat and cool the plate 102. The detecting means preferably includes a detector 108 connected to the chambers which produces a chamber signal for each chamber corresponding to genetic material in each chamber. The detecting means preferably also includes a processor 110 in communication with the detector 108 which receives the signal and identifies unique properties of the nucleotides in each chamber. The unique properties of the nucleotide of the genetic material in each chamber pertain to a number of nucleotides in any of the nucleotide sequences of the genetic material.

The amplifying means preferably includes at least one nucleotide sequence that corresponds to each chamber and which is in contact with the chamber. Each nucleotide sequence interacts with the nucleotide sequence of the genetic material of the nucleotide sequence if it is present.

The present invention also pertains to a method for analyzing genetic material of an organism. The method comprises the steps of amplifying the genetic material. Then there is the step of characterizing the amplified genetic material in a region having a radius of less than 20 feet at a rate exceeding 100 sequence-tagged sites per hour per organism. Preferably, the genetic material includes RNA or DNA. After the characterizing step, there preferably is th step of accessing risk of illness for which there is a genetic susceptibility in the organism. Such illnesses can include cancer, heart disease, etc.

The present invention also pertains to a method for manufacturing an apparatus for analyzing genetic material of an organism. The method comprises the steps of placing corresponding sequence-tagged sites in contact with corresponding chambers of a plate 102. Then, there is the step of connecting detectors 108 to the chambers which can detect where the nucleotide sequences of the genetic material of the organism, when placed in contact with the chambers, have reacted with the corresponding sequence-tagged sites in the corresponding chamber. Then, there is the step of placing a thermocycling device 104 in contact with the plate 102 to cause the sequence-tagged sites in the chambers to react with genetic material of the organism that is placed in contact with the chambers. Next, there is the step of connecting a computer 110 to the detectors 108 and to the

thermocycling device 104 to control operation of the thermocycling device 104, and to receive signals which correspond to the genetic material of the organism and the sequence-tagged sites of each chamber from the detectors 108.

The present invention also pertains to a method for determining the size of nucleotide sequences of an STR marker contained on genetic material comprising the steps of: amplifying the nucleotide sequences of the genetic material in a region relating to the STR marker. Then there is the step of performing nucleic acid hybridizations on the amplified nucleotide sequences. Then there is the step of producing signals corresponding to the hybridizations of the amplified nucleotide sequences. Then there is the step of determining the sizes of the nucleotide sequences contained in the genetic material.

A. An Apparatus for Parallel Genotyping

A parallel genotyping apparatus is described. The purpose of said apparatus is to provide a physical, chemical, mechanical, and computational embodiment for performing simultaneous experiments on multiple genetic markers used for genetic characterization.

Referring to figure 1, the apparatus is comprised of the following components:

(1) A multi-chambered reaction plate 102.
(2) A thermocycling device 104.
(3) A robotic device 106.
(4) A det ction device 108.
(5) A computer device 110, with a memory.

The biochemical reactions occur in the chambers of the reaction plate 102, wherein a "chamber" denotes any localized region suitable for performing said reactions. The thermocycling device 104 provides a means for PCR and hybridization experiments. The robotic device 106 provides a means for transferring chemicals and performing other physical/chemical operations. The detection device 108 is used to quantitatively measure the signals from DNA hybridization experiments. The computer device 110 coordinates the activity of the other components, and performs any needed computations.

The primary requirement of the multi-chambered reaction plate 102 is a set of spatially arrayed chambers, each containing its own PCR primers for genome characterization, and providing operations for PCR amplification, DNA hybridization, and signal detection. Any physical device, of any number of dimensions, in whole or in part, that provides this functionality can serve as a physical embodiment for the apparatus. In an alternative embodiment, parallel synthesis methods for producing the oligonucleotides by spatially addressable masking techniques on a surface have been described (Fodor, S.P.A., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., and Solas, D. 1991. Light-directed spatially addressable parallel chemical synthesis. *Science*, 251: 767-773), incorporated by reference, and may be employed for manufacture. The process may be further miniaturized using molded or etched surfaces that allow one or more orders of magnitude of markers to be simultaneously characterized in each chamber without increasing DNA or enzyme requirements.

In the preferred embodiment, the basic container for the parallel genotyping reactions is a commercially available polystyrene or polycarbonate 384-chamber microtiter plate (USA Scientific Products, Ocala, Fl). Alternative embodiments include 96-chamber and 864-chamber plates. Each chamber corresponds to one chamber. These plates occupy the space of standard 96-chamber microtiter plates and are compatible with current robotic systems such as the Beckman Biomek system. These plates can contain sufficient volumes for the PCR reactions in each chamber. Many of the required mechanical, physical, and chemical steps can be performed on the plate by manipulating it with currently available robotic units (e.g., Beckman Biomek) (Bentley, D.R., Todd, C., Collins, J., Holland, J., Dunham, I., Hassock, S., Bankier, A., and Giannelli, F. (1992). The development and application of automated gridding for efficient screening of yeast and bacterial ordered libraries. *Genomics*, 12(3): 534-41. Civitello, A.B., Richards, S., and Gibbs, R.A. (1992). A simple protocol for the automation of DNA cycle sequencing reactions and polymerase chain reactions. *Dna Sequence*, 3(1): 17-23. Drmanac, R., Drmanac, S., Labat, I., Crkvenjakov, R., Vicentic, A., and Gemmell, A. (1992). Sequencing by hybridization: towards an automated sequencing of one million M13 clones arrayed on membranes. *Electrophoresis*, 13(8): 566-73.), incorporated by reference, as described below.

The apparatus has one or more two-dimensional surfaces 102 comprised of reaction chambers. Each STS genetic marker used from a genome corresponds to some reaction chamber. This experimentation surface provides a means for performing parallel laboratory operations on all the chambers simultaneously. Within each chamber, five steps are performed:

(1) A deposition of at least two oligonucleotides into the chamber. These oligonucleotides serve as PCR primers for the STS marker specific to the chamber.

(2) A PCR amplification of genomic DNA presented to the chamber.

(3) A DNA hybridization experiment that characterizes the amplified DNA, and possibly modifies the DNA.

(4) A signal detection from the hybridized (and possibly modified) DNA.

(5) An analysis of the detected signals to determine the alleles of the specific STS marker.

Means are provided by the apparatus for PCR amplification, DNA hybridization, and signal detection. The following description relates these functions to the parts of the apparatus.

Deposit primers. This function can be considered part of the manufacturing process, as described below.

PCR Amplification. The apparatus provides the means for amplifying the STS DNA region subsequent to presentation with genomic DNA. When PCR (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. 1990. *PCR Protocols: A Guide to Methods and Applications*. San Diego, CA: Academic Press. Mullis, K.B., Faloona, F.A., Scharf, S.J., Saiki, R.K., Horn, G.T., and Erlich, H.A. 1986. Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.*, 51: 263-273.), incorporated by reference, is used for the amplification step, this means includes thermocycling components for heating and cooling the reaction mixture. In the preferred embodiment, the genomic DNA and PCR reagents are simultaneously transferred to the chambers by means of

the robotic device. Various thermostable polyermases can be used (Garrity, P.A., and Wold, B.J. (1992). Effects of different DNA polymerases in ligation-mediated PCR: enhanced genomic sequencing and in vivo footprinting. *Proceedings of the National Academy of Sciences of the United States of America*, **89**(3): 1021-5. Ling, L.L., Keohavong, P., Dias, C., and Thilly, W.G. (1991). Optimization of the polymerase chain reaction with regard to fidelity: modified T7, Taq, and vent DNA polymerases. *Pcr Methods & Applications*, **1**(1): 63-9.), incorporated by reference.

In the preferred embodiment, thermocycling is done using a conventional programmable block thermal cycler 104 based on the heating and cooling of a metal block (using Peltier or fluid refrigerants for cooling) (R. Hoelzel, Trends in Genetics, August 1990, volume 6 #8; p 237-8), incorporated by reference. The reaction plate is transferred to and from this computer-controlled thermal cycler by means of the robot 106. In an alternative embodiment, a device 104 is used that heats and cools a rapidly circulating air mass around the plate (e.g., Biotherm PCR oven) (Garner, H.R., Armstrong, B., and Lininger, D.M. (1993). High-throughput PCR. *Biotechniques*, **14**(1): 112-5.), incorporated by reference. Such air thermal cyclers support the simultaneous processing of multiple plates. The conditions (such as temperature settings and ramp functions and step times) are adjusted to the method of heat and cooling, since the sensitivity of the method to how rapidly the reaction chambers will equilibrate with the changing temperatures.

In an alternative embodiment, a robotic attachment (Beckman Biomek), incorporated by reference, comprised of a thermocycling surface which has the same 384-chamber shape as

the reaction plate is used to physically mate with the 384-chamber reaction plate, and provide the necessary heating and cooling operations under computer control. In another alternative embodiment where the reaction surface is fabricated, heating and cooling elements such as Peltier junctions can be physically incorporated into the apparatus. This surface is suitable for transferring sample genomic DNA to many chambers simultaneously. Miniaturization enable shorter cycle times and greater homogeneity because of the rapid temperature equilibration of the thin films and small volumes.

DNA hybridization. Sufficient volume and chemical composition is provided within each reaction chamber so that the requisite DNA hybridization (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., ed. 1993. *Current Protocols in Molecular Biology*. New York, NY: John Wiley and Sons. Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning, second edition*. Plainview, NY: Cold Spring Harbor Press.), incorporated by reference, can occur. In the preferred embodiment, the robotic component of the apparatus transfers the hybridization reaction mixture to the chambers, and provides means for heating and cooling the reaction chamber, as described above.

In a preferred embodiment, means are provided for (optionally) modifying the DNA. Typical modifications to heteroduplexes, for example, include chemical derivatization and endonuclease digestion of single-stranded components.

Signal Detection. The detection of the heteroduplexes and nucleotides within the loops is done with

a commercially available spectrophotometric/fluorometric instrument 108 similar to that used for ELISAs (Dynatech Laboratories, Chantilly, Va), incorporated by reference, modified to accomodate the larger number and smaller size chambers. A scanning laser fluorimeter can also be employed over the plate surface. Because the plate is flat and comprised of an optical grade surface, fluorescent detection is straightforward. The robot transfers the reaction plate to this optical detection device prior to the detection operation. In an alternative embodiment, computerized fluorescent scanning microscopes are used that are capable of detecting and quantitating fluorescent signals and ar suitable for the miniaturized system. These have been developed for immunological and genetic cytochemistry (Biological Detection Systems), incorporated by reference.

A physical signal is measured from the reagent attached to a PCR primer. In other alternative embodiments, such detection reagents include (but are not limited to) radioactivity, fluorescence, phosphorescence, chemiluminescence, electrical resistivity, pH, and ionic concentration. The direct electrical detection mechanisms are particularly attractive for direct coupling of the experiment onto a minaturized solid state detection device (Briggs, J., Kung, V.T., Gomez, B., Kasper, K.C., Nagainis, P.A., Masino, R.S., Rice, L.S., Zuk, R.F., and Ghazarossian, V.E. (1990). Sub-femtomole quantitation of proteins with Threshold, for the biopharmaceutical industry. *Biotechniques*, 9(5): 598-606. Kung, V.T., Panfili, P.R., Sheldon, E.L., King, R.S., Nagainis, P.A., Gomez, B.J., Ross, D.A., Briggs, J., and Zuk, R.F. (1990). Picogram quantitation of total DNA using DNA-binding proteins in a silicon sensor-based system. *Analytical Biochemistry*, 187(2): 220-7. Olson, J.D., Panfili,

P.R., Armenta, R., Femmel, M.B., Merrick, H., Gumperz, J., Goltz, M., and Zuk, R.F. (1990). A silicon sensor-based filtration immunoassay using biotin-mediated capture. *Journal of Immunological Methods*, **134**(1): 71-9. Olson, J.D., Panfili, P.R., Zuk, R.F., and Sheldon, E.L. (1991). Quantitation of DNA hybridization in a silicon sensor-based system: application to PCR. *Molecular & Cellular Probes*, **5**(5): 351-8.), incorporated by reference. Such silicon-based detectors are described below.

Analysis. The analysis of the signals is done by a computer device 110. Means are provided for the signals are transferred from the detector into the memory of the computer. A computer program for determining genotypes from the quantitative signals and calibrations curves resides in the memory of said computer.

B. Manufacturing an Apparatus for Parallel Genotyping

In the preferred embodiment, the apparatus is manufactured by selecting a set of genetic markers, synthesizing both standard and derivatized oligonucleotide primers, and then depositing said oligonucleotide primers into the reaction chambers of a 384-chamber plate. This plate is then positioned with the other components of the apparatus, including the thermocycling device, the robotic device, the detection device, and the computer device.

A sufficient number of polymorphic genetic markers are chosen for unambiguously characterizing or tracing chromosomes in an organism containing DNA or RNA. Depending on the application, this can range from 10 centiMorgan (cm) to 0.001 cm. One cm is approximately one million megabases

(Mb). In a preferred embodiment, a resolution of 0.1 cm, or 100,000 base pairs (bp), is used. In the human species, for example, which contains about 3 billion bp, this works out to 30,000 markers. The genetic markers to be used for each STS are obtained as PCR primer sequences pairs from available databases (Genbank, GDB, EMBL; Hilliard, Davison, Doolittle, and Roderick, Jackson laboratory mouse genome database, Bar Harbor, ME; SSLP genetic map of the mouse, Map Pairs, Research Genetics, Huntsville, AL), incorporated by reference. One of the goals of the world wide genome project is to generate and make publicly available 30,000 genetic markers; currently, about 10,000 are available. Alternatively, some or all of these PCR sequences can also be constructed using existing techniques (Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning, second edition*. Plainview, NY: Cold Spring Harbor Press.), incorporated by reference.

The oligonucleotide primers for each STS are synthesized (Haralambidis, J., Duncan, L., Angus, K., and Tregear, G.W. 1990. The synthesis of polyamide-oligonucleotide conjugate molecules. *Nucleic Acids Research*, 18(3): 493-9. Nelson, P.S., Kent, M., and Muthini, S. 1992. Oligonucleotide labeling methods. 3. Direct labeling of oligonucleotides employing a novel, non-nucleosidic, 2-aminobutyl-1,3-propanediol backbone. *Nucleic Acids Research*, 20(23): 6253-9. Roget, A., Bazin, H., and Teoule, R. 1989. Synthesis and use of labelled nucleoside phosphoramidite building blocks bearing a reporter group: biotinyl, dinitrophenyl, pyrenyl and dansyl. *Nucleic Acids Research*, 17(19): 7643-51. Schubert, F., Cech, D., Reinhardt, R., and Wiesner, P. 1992. Fluorescent labelling of sequencing primers for automated oligonucleotide synthesis. *Dna Sequence*, 2(5):

273-9. Theisen, P., McCollum, C., and Andrus, A. 1992. Fluorescent dye phosphoramidite labelling of oligonucleotides. *Nucleic Acids Symposium Series*, 1992(27): 99-100.), incorporated by reference. These primers may be derivatized with a fluorescent detection molecule or a ligand for immunochemical detection such as digoxigenin. Derivatization of the primer for binding to the surface entails the incorporation of a biotinylated nucleotide at the 5' end of the synthetically made oligonucleotide. Additional biotinylated residues can also be incorporated (depending on the protocol) into this primer either at the time of biosynthesis or by secondary photo or chemical biotinylation. Though the preferred embodiment employs the direct addition of the 5' biotin by chemical synthesis, additional biotin molecules for binding may be added to the primer for improving the efficiency of selection of heteroduplexes for analyses. Alternatively, said oligonucleotides and their derivatives can be ordered from a commercial vendor (Research Genetics, Huntsville, AL).

The oligonucleotide primer sets are deposited into each reaction chamber by means of a robotic system from source chambers containing a large store of presynthesized oligonucleotides. Said transferring can be effected in one or more operations, wherein oligonucleotide primers are deposited into multiple chambers in each transferring step, thereby creating a two-dimensional spatial array. In an alternative embodiment, this deposition is effected by means of a parallel deposition device to which the 384-chamber plate is presented by means of a conveyor belt. The deposition device has source chambers, each containing a large store of a unique oligonucleotides specific to a reaction chamber. Said source chambers are spatially arrayed

to conform to the reaction chambers of the plate. Both the device and plate are properly positioned and made stationary, and then the chambers are filled in one or more more steps with the oligonucleotide.

In the preferred embodiment, the plates are dried and each chamber is then coated with a wax material, such as Ampliwax (Perkin-Elmer, Norwalk, CT), incorporated by reference. This material hardens at 4°C, is liquid throughout the temperature range of the PCR, and serves as a vapor barrier to prevent evaporation of the PCR reactions during the denaturation steps at 95°C. By placing this material over the dried primers and allowing it to harden at 4°C, one establishes a stable apparatus that can be stored and to which the remaining components of the PCR reaction can be added without disruption of the stable two-dimensional array and the reactions can be initiated simultaneously.

In an alternative embodiment, the oligonucleotides are covalently attached to a substrate such as glass by spatially addressable light-directed parallel DNA synthesis (Drmanac, R., Drmanac, S., Strezoska, Z., Paunesku, T., Labat, I., Zeremski, M., Snoddy, J., Funkhouser, W.K., Koop, B., and Hood, L. 1993. DNA Sequence Determination by Hybridization: a Strategy for Efficient Large-Scale Sequencing. *Science*, 260: 1649-1652. Fodor, S.P.A., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., and Solas, D. 1991. Light-directed spatially addressable parallel chemical synthesis. *Science*, 251: 767-773.), incorporated by reference. The DNA amplification is done directly on this surface (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. 1990. *PCR Protocols: A Guide to Methods and Applications*. San Diego, CA: Academic Press.), incorporated

by reference. Detection can be effected by ^{32}P, fluorescence, or electronic means (Eggers, M., et. al., 1993. Genosensors: Microfabricated Devices for Automated DNA Sequence Analysis. In *Advances in DNA Sequencing Technology*, #1891. Keller, R., ed., Proceedings of SPIE. Southern, E.M., Maskos, U., and Elder, J.K. 1991. Analyzing and Comparing Nucleic Acid Sequences by Hybridization to Arrays of Oligonucletides: Evaluation Using Experimental Models. *Genomics*, 13: 1008-10017.), incorporated by reference.

Other components of the apparatus include resins and filters that will nonspecifically and reversibly bind double-stranded DNA, but not free nucleotides or short oligonucleotides (Molecular Biology LabFax, T.A. Brown, ed. Academic Press p281-4), incorporated by reference. These are commercially available and can be readily modified to be fit within a manifold that will ensure leak-proof contact with the reaction chambers or plates. Uncharged nylon, charged nylon, and nitrocellulose are some of the filter materials in current use (Harley, C.B., and Vaziri, H. (1991). Deproteination of nucleic acids by filtration through a hydrophobic membrane. *Genetic Analysis, Techniques & Applications*, 8(4): 124-8. Twomey, T.A., and Krawetz, S.A. (1990). Parameters affecting hybridization of nucleic acids blotted onto nylon or nitrocellulose membranes. *Biotechniques*, 8(5): 478-82. Williams, D.L. (1990). The use of a PVDF membrane in the rapid immobilization of genomic DNA for dot-blot hybridization analysis. *Biotechniques*, 8(1): 14-5.), incorporated by reference. The polystyrene plate that is bound to strepavidin (or avidin) is also commercially available in neutral, positively-, and negatively-charged configurations (MaxiSorp, Nunc, or Combiplate, Applied

Scientific Instrumentation, Inc, Eurgene, OR), incorporated by reference. The appropriate material is adjusted to the specific combination of the binding capacity, the degree of nonspecific or background binding, and the optical properties of the material.

In the preferred embodiment, referring to figure 1, the commercially available polystyrene or polycarbonate 384-chamber microtiter plate 102 is arranged in a 24 by 16 array. The commercially available robotic device 106 has a surface with 384 chambers arranged in a spatial configuration identical to that of the reaction plate 102. Thus, all robotic actions (e.g., for the steps of amplification, hybridization, and detection) are performed in parallel with robotic device 106 in mechanical juxtaposition with plate 102.

In the preferred embodiment, the commercially available programmable block thermal cycler 104 has a surface with 384 chambers arranged in a spatial configuration identical to that of the reaction plate 102. During thermocycling, every chamber of the plate 102 is in direct contact with its corresponding chamber in the thermocylcer 104. In an alterative embodiment, the commercially available programmable oven thermocyler 104 is sufficiently large to accommodate the dimensions of 384-chamber reaction plate 102, and has sufficient uniformity to perform the necessary amplification reactions within each chamber. A robotic device is used to transfer the reaction plate 102 to and from the oven thermocycler 104.

The commercially available ELISA-like spectrophotometric/fluorometric detection device 108 contains

384 chambers arranged in an spatial configuration identical to that of reaction plate 102. During the detection phase, the plate 102 is placed into the detector, with each chamber of plate 102 residing within its corresponding detection chamber of detector 108. This enables detections to be conducted simultaneously and independently for each chamber.

The computer device 110 coordinates the activities of the other components plate 102, thermocyler 104, robotic device 106, and detector 108. Note that most commercial thermocylers, robotic devices, and detectors include computational facilities for independently performing control, detection, and processing tasks, thus freeing the computer device 110 from such low-level processes. The computer device 110 is connected to the detector 108. Signals obtained from the detector 108 are transferred to the memory of computer 110. The computer 110 employs processing means for interpreting the signals in its memory, and determines and outputs the characteristics of nucleotid sequences in each chamber of the reaction plate 102.

C. A System for Parallel Genotyping

A system for characterizing multiple genetic markers is described, along with steps for using this information for preventative health care. In overview of the preferred embodiment, genomic DNA is first extracted from an individual (say, by processing a blood sample). PCR reagents are then mixed with the genomic DNA, and a robotic device applies this PCR/DNA mixture to the chambers of the reaction plate of the apparatus. Every chamber has its own predeposited PCR primers that define a unique genetic marker. PCR amplificiation of the genomic DNA marker region is then

performed on every chamber using the thermocycling component of the apparatus. A quantitative hybridization experiment is then conducted in every chamber, possibly modifying the DNA. The signals from these hybridization experiments are then measured from every chamber using the detection component, such as fluorescence measurements with a scanning light microscope. More than one (e.g., two or three) such parallel experiments may be needed to acquire all necessary genotyping data for one STR. The measurements are then collected and analyzed by the component computer device to characterize the alleles at every marker.

The resulting genotyping information from the multiple alleles can be used for a number of applications, as described below. One important use is the determination of genetic risk for phenotypic traits, including diseases. By comparing dense genotyping data of STRs across related individuals, haplotypes can be compared, and the shared genomic regions determined. Correlating a shared trait and genotype commonalities enables a determination of genomic patterns that imply a quantitative risk for said trait. These patterns can be applied to the genotypes of an individual and their relatives to compute a probability of expressing the trait. When the traits correspond to common multigenic multifactorial diseases, the highest risk entities are determined, and preventative measures undertaken, thereby improving the health of said individual. Software systems are built to tailor the genotyping information for this advising task.

The quantitative hybridization experiment that is used in the preferred embodiment is a pair of loop mismatch assays. The first assay measures the sum of the two STR

allele loops, relative to a third (and smaller) STR. The second assay measures the difference of the two STR allele loops relative to each other. By combining the sum and difference values, the two alleles can be determined. The quantitative loop detection is effected by directly measuring the signals derived from the loops relative to the number of strands with loops (this is described in detailed later on). The loops are quantitated either by a chemical modification of the single-stranded loop DNA into a detectable state, or by incorporation of labeled DNA and subsequent digestion and detection of the single-stranded loop. The number of strands is measured by using an end-labeled PCR primer. The ratio of the (calibrated) loop measurements to the number of strands determines the loop size. In an alternative embodiment, multiple hybridizations are performed for every STR, producing a patter that determines the genotype.

This system for performing multiple genotypings in parallel, with each STR in its separate cell, has many useful advantages over current genotyping methods, including the best gel-based multiplex methods. Specifically,

- Massive parallelism greatly increases throughput by greatly reducing the total experimentation time.

- The experiment's architecture allows independent interchangeability of STR loci. Any STR(s) of the same class can be placed at any cell of the device.

- The synthesis of oligonucleotides can be spatially or temporally separated from the execution of the PCR amplification and the detection.

- Manufacturing enables miniaturization of the device, and the incorporation of detection machinery into the device.

- The manual labor required for genotyping is greatly reduced, because the manufactured device eliminates the separate steps of handling multiple (e.g., thousands) specific STR primers. This includes synthesizing the oligonucleotides, performing the PCR, loading gels or other detection devices, and checking the genotyping results.

- Reduced manual intervention greatly reduces the error rate.

Referring to figure 2, the following terminology is used throughout:

Strand. A single-stranded DNA PCR product of an STR. The CA (or GT) repeat region is of varying length.

Complementary strand. A second strand having a Watson-Crick complementary DNA sequence to a first strand. However, the number of CA or GT repeats need not equal that of the first strand.

Upper strand. The DNA strand 202 of the STR locus that contains the CA-repeat units.

Lower strand. The DNA strand 204 complementary to the upper strand that contains the GT-repeat units.

Left primer. The PCR oligonucleotide primer 206 that initiates the upper strand of the STR locus.

Right primer. The PCR oligonucleotide primer 208 that initiates the lower strand of the STR locus.

In the preferred embodiment, the system is comprised of the following steps:

Referring to figure 3, Step 1 entails the manufacture of an apparatus in which STR loci have been selected, and appropriate oligonucleotides (with modifications) synthesized and deposited within each chamber.

In the PCR ampification of Step 2a, the process begins by extracting DNA from blood or tissue. There are numerous standard methods to isolate DNA including whole blood, isolated lymphocytes, tissue, and tissue culture (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., ed. 1993. *Current Protocols in Molecular Biology*. New York, NY: John Wiley and Sons. Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning, second edition*. Plainview, NY: Cold Spring Harbor Press. Nordvag 1992. Direct PCR of Washed Blood Cells. *BioTechniques*, 12(4): 490-492.), incorporated by reference. In the preferred embodiment, DNA is extracted from anticoagulated human blood removed by standard venipuncture and collected in tubes containing either EDTA or sodium citrate. The red cells are lysed by a gentle detergent and the leukocyte nuclei are pelleted and washed with the lysis

buffer. The nuclei are then resuspended in a standard phosphate buffered saline (pH=7.5) and then lysed in a solution of sodium dodecyl sulfate, EDTA and tris buffer pH 8.0 in the presence of proteinase K 100 ug/m l. The proteinase K digestion is performed for 2 hours to overnight at 50°C. The solution is then extracted with an equal volume of buffered phenol-chloroform. The upper phase is reextracted with chloroform and the DNA is precipitated by the addition of NaAcetate pH 6.5 to a final concentration of 0.3M and one volume of isopropanol. The precipitated DNA is spun in a desktop centrifuge at approximately 15,000 g, washed with 70% ethanol, partially dried and resuspended in TE (10mM Tris pH 7.5, 1 mM EDTA) buffer. There are numerous other methods for isolating eucaryotic DNA, including methods that do not require organic solvents, and purification by adsorption to column matrices. None of these methods are novel, and the only requirement is that the DNA be of sufficient purity to serve as templates in PCR reactions and the amount of DNA is sufficient for the scale of the parallel genotyping procedure.

Continuing Step 2a, the reaction plates of the apparatus are maintained at 4°C at the time the genomic DNA has been mixed with the other components of the PCR reaction. These other components include, but are not limited to, the standard PCR buffer (containing Tris pH8.0, 50 mM KCl, 2.5 mM magnesium chloride, albumin), triphosphate deoxynucleotides (dTTP, dCTP, dATP, dGTP), the thermostable polymerase (Taq polymerase in this preferred embodiment, but others are available though buffer conditions are somewhat different) (Garrity, P.A., and Wold, B.J. 1992. Effects of different DNA polymerases in ligation-mediated PCR: enhanced genomic sequencing and in vivo footprinting. *Proceedings of the*

National Academy of Sciences of the United States of America, 89(3): 1021-5. Ling, L.L., Keohavong, P., Dias, C., and Thilly, W.G. 1991. Optimization of the polymerase chain reaction with regard to fidelity: modified T7, Taq, and Vent DNA polymerases. *PCR Methods & Applications*, 1(1): 63-9.), incorporated by reference. The amounts of the genomic DNA and deoxynucleotides, Mg concentration, and enzyme are all adjusted so as to be optimal for the entire set of PCR reactions. The PCR primers for each locus are chosen for consistency with these uniform reaction conditions. The total amount of this mixture is determined by the final volume of each PCR reaction (say, 10 ul) and the number of reactions (say, 384). This mixture can also be varied by including some of the constituents with the primers that are previously deposited in the microchambers. All of the necessary components for the PCR reactions are kept separate until the Ampliwax is melted and the aqueous phases reconstitute, each reaction cell receives a consistent and reproducible amount of the necessary components, and the combination of constituents does not compromise stability and biological activity (e.g., the Taq polymerase may be unstable if stored in a lyophilized state on the reaction plates).

In Step 2b, the DNA/PCR mixture is applied to the reaction chambers with the Biomek robotics unit and the PCR is initiated by heating the plate rapidly to 95°C in order to melt the ampliwax, allow the DNA/PCR mixture to mix with the oligonucleotide primers (convection mixture is sufficent), and denature the genomic DNA. The ampliwax forms a stable vapor barrier over the chambers during the PCR reactions. This method of initiating the PCR reactions is referred to as a "hot start" (D'Aquila et al., Nuc. Acid Res. 19 (13) 3749 (1991)), incorporated by reference, and has the additional

benefit of reducing the amount of nonspecific PCR products that are produced, thus improving the purity and amount of the final desired PCR signal that will be detected.

In Step 2c, the PCR reactions are performed on all of the reactions simultaneously by appropriately heating and cooling the plate to specific temperatures. After the initial denaturation step of 93°-95°C for 3-5 minutes, the plates are cooled to the annealing temperature (50°-65°C, typically 55°C) for a set time (0-100 seconds, typically 15 seconds), warmed to the extension temperature which is optimal for the thermostable polymerase (e.g., 73°C for Taq polymerase) and maintained for a set period of time (0-100 seconds, typically 30 seconds). Finally, the cycle is completed by elevating the temperature of the reaction to denature the DNA products (93-95°C for 0 - 60 seconds, typically 15 seconds). The entire cycle of annealing, extension, and denaturation is then repeated multiple times (ranging from 20-40 cycles depending on the efficiencies of the reactions and sensitivity of the detection system). (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. 1990. *PCR Protocols: A Guide to Methods and Applications*. San Diego, CA: Academic Press.), incorporated by reference.

Following Step 2c, the PCR cycles are completed, with each chamber containing the amplified DNA from a specific location of the genome. Each mixture includes the DNA that was synthesized from the two alleles of the diploid genome (a single allele from haploid chromosomes as is the case with the sex chromosomes in males or in instances of cells in which a portion of the chromosome has been lost such as occurs in tumors, or no alleles when both are lost). Also in

this mixture are the free triphosphate deoxynucleotides and the unused oligonucleotide primers.

Step 2d is the last PCR step, which inactivates the thermostable polymerase, say, by the addition of EDTA. Ampliwax protects the integrity of the chambers and the mixing occurs at 37°C for several minutes.

In Step 3a, for quantitive loop mismatch genotyping, the DNA strands are allowed to reanneal at a temperature above the annealing temperature of the oligonucleotide primers, but below the melting temperature of perfectly matched complementary strands. In most instances, this will be between 65 and 75°C, depending on the salt conditions of the buffer. The annealing time can vary from 1 hour to 24 hours, with 2 hours selected in the preferred embodiment.

In Step 3a, the heteroduplex annealing is done with the original contents of the chamber for the "subtraction" assay of the loop detection method. The "addition" assay that is required for the measurements of loop mismatches entails combining of the contents of a chamber with its counterpart from a control plate in which the PCR reaction has been carried out with a corresponding set of primers (same oligonucleotides, but with different primer modifications) on a target DNA that has the smallest possible number of repeated elements for the given DNA marker. These two assays are done in different chambers of the reaction plate, or on separate plates entirely.

In the subtraction assay, the Left primer is linked to a detection molecule, and the Right primer is covalently linked to a molecule necessary for binding (i.e., biotin in the preferred embodiment).

In the addition assay, the unknown genomic DNA (Source DNA) is amplified using a Left primer that is labeled with the detection molecule and the Right primer is unmodified. In contrast, the control DNA (or Target DNA) is amplified with an unmodified Left primer and the Right primer contains the binding protein (such as biotin). In this situation, when the amplified DNA from the unknown source and the Target DNA are combined to form heteroduplexes, one will only detect the binding of the upper strand of the Source DNA to the immobilized lower strand of the Target DNA and homoduplexes of the Target DNA strands will be undetected as well as perfectly matched, creating no exposed loops for detection. The corresponding Source and Target DNAs are appropriately combined using the Biomek robot though direct physical transfer methods (i.e., aligning the Source DNA plate on top of the Target DNA plate directly and mixing by melting th ampliwax).

In Steps 3b, 3c, 3d, 3e, and 3f, the unwanted singl strands, primers and free nucleotides are removed by using a 3'to5'-specific exonuclease that will not cleave or disrupt internal single-stranded loop structures, in both the subtraction and addition assays. Exonuclease VII from E. coli is capable of 3'- 5' exonuclease activity limited to single-stranded DNA (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., ed. 1993. *Current Protocols in Molecular Biology*. New York, NY: John Wiley and Sons. Vales, L.D., Rabin, B.A., and Chase, J.W. 1982. Subunit structure of Escherichia coli exonuclease VII. *J. Biol. Chem.*, 257: 8799-8805.), incorporated by reference. One unique feature of this exonuclease is that it is not inactivated by EDTA, thus making it active under conditions that would inactivate the

Taq polymerase. The enzyme is either removed by a subsequent wash step, or inactivated chemically. (A 5' exonuclease is not used, because that end is blocked by the linkage t biotin or by linkage to the surface.) The enzyme is added to the chambers over the Ampliwax surface, allowed to mix at 37°C and incubated for a brief period (1-60 minutes, 10 minutes in the preferred embodiment) and terminated by the addition of EDTA. At the same time, the buffer is adjusted to promote non-specific binding of the DNA to a resin or filter.

Following Step 3f, free deoxynucleotides and primers that interfere with binding of the PCR products and the detection system are removed. The purification of unincorporated DNA materials is combined with the elimination of single-stranded DNA species that remain after heteroduplex formation. In the preferred embodiment, this purification step is done after the heteroduplex formation, thereby also eliminating single-stranded DNA's. Although heteroduplex formation may be somewhat inhibited by residual primers, combining of the steps greatly simplifies the method and aids in increasing the signal-to-noise ratio. In an alternative embodiment, the separation of free deoxynucleotides and primers from the PCR products is achieved by filtration (the unwanted materials are significantly smaller than the final PCR products) using commercially available filters (Centricon 30 filters, Amicon), incorporated by reference, or by adsorption (Molecular Biology LabFax, T.A. Brown, ed. Academic Press p281-4), incorporated by reference, which entails the nonspecific binding of the PCR products, double-stranded and single-stranded DNA to a matrix followed by removal of the supernatents containing the primers and nucleotides. The removal of the primers, followed by

heteroduplex formation and then elimination of single strands can be done by exonuclease digestion or by other separation methods (Linxweiler, W., and Horz, W. 1982. Sequenc specificity of exonuclease III from E. coli. *Nucleic Acids Research*, 10(16): 4845-59. Sandigursky, M., and Franklin, W.A. 1993. Exonuclease I of Escherichia coli removes phosphoglycolate 3'-end groups from DNA. *Radiation Research*, 135(2): 229-33.), incorporated by reference.

In Step 3g, the filter is set upon a plastic manifold that fits over the chambers of the 384 chamber plate, the apparatus is inverted so that the Ampliwax rises to the bottom surface of the chambers, and the DNA solution comes into contact with the filter. In Step 3h, the filter is separated from the chamber and washed with a high salt buffer to remove the free nucleotides. In Step 3i, this filter is then placed against a polystyrene surface (optical grade) (such as used in MaxiSorp plates manufactured by Nunc, Naperville, IL) that has been coated with streptavidin (Giorda, R., Lampasona, V., Kocova, M., and Trucco, M. 1993. Non-Radioisotopic Typing of Human Leukocyte Antigen Class II Genes on Microplates. *BioTechniques*, 15(5): 918-925. Giorda et al. 1994. Molecular HLA DQ typing on microplates: A step toward complete automation. manuscript), incorporated by reference and the DNA's are eluted from the filter using a low ionic strength buffer such as TE (10 mM Tris pH7.5, 1 mM EDTA), and allowed to specifically bind to the streptavidin through the biotinylated primers described previously.

The heteroduplexes are bound to the polystyrene surface in an exact replica of their initial spatial orientation. The heteroduplexes containing biotinylated primer will bind to the streptavidin surface under a wide range of buffers that

are pH neutral. In the preferred embodiment, the DNA is bound in the TE buffer and in Step 3j the plate is washed twice with 0.15 M phosphate buffer.

In Step 3k, the chemical derivatization of the C and A residues within the heteroduplex loops employs a modification of the method originally described by (Kimura, K., Nakanishi, M., Yamamoto, T., and Tsuboi, M. (1977). A correlation between the secondary structure of DNA and the reactivity of adenine residues with chloroacetaldehyde. *Journal of Biochemistry*, 81(6): 1699-703.), incorporated by reference. The plate is washed with 0.15 M Na Phosphate buffer pH = 6.5 (the pH can be varied from 4.5 to 6.5 and alternative buffers can be used). The plate is then covered with the buffer containing a final concentration of CAA is 2.0% and incubated at 37°C for 4 hours (longer or shorter times may be used). The reaction is terminated in Step 31 by washing with 0.01M Tris-HCl pH 7.0 and 1.0 M NaCl. The NaCl prevents dissociation of the heteroduplexes during the etheno-dehydration step. The plate is heated in the final wash volume at 85-90°C for 1 hour, which dehydrates the ethenoderivative. Note that loop-specific derivatization of the nucleotides with chloracetaldehyde or other chemical modification reagents (osmium tetroxide, hydroxylamine, carbodiimide, etc., as described below) provides an alternative means for eliminating background reagents prior to detecting nuclease-liberated free derivatized nucleotides.

In the Step 4a detection, the fluorescence of the primer detector molecule that is bound to the hybridized strand is measured at this time, or measured at a later stage in conjunction with the fluorescent adducts created within the loop structures. In the preferred embodiment, the detection

of the hybridized strands and the derivatized nucleotid s within the loops are performed at the same time. The method of detection is preferrably by fluorescence (Kimura, K., Nakanishi, M., Yamamoto, T., and Tsuboi, M. (1977). A correlation between the secondary structure of DNA and the reactivity of adenine residues with chloroacetaldehyde. *Journal of Biochemistry*, **81**(6): 1699-703.), incorporated by reference. Alternative embodiments include chemiluminesence (Martin R., Hoover, C., Grimme, S., Grogan, Cl, Holtke, J. and Kessler, CF. (1990) Bio Techniques 9(6): 762-8), incorporated by reference, electrochemical coupling using silicon surfaces (Briggs, J., Kung, V.T., Gomez, B., Kasper, K.C., Nagainis, P.A., Masino, R.S., Rice, L.S., Zuk, R.F., and Ghazarossian, V.E. (1990). Sub-femtomole quantitation of proteins with Threshold, for the biopharmaceutical industry. *Biotechniques*, **9**(5): 598-606. Kung, V.T., Panfili, P.R., Sheldon, E.L., King, R.S., Nagainis, P.A., Gomez, B.J., Ross, D.A., Briggs, J., and Zuk, R.F. (1990). Picogram quantitation of total DNA using DNA-binding proteins in a silicon sensor-based system. *Analytical Biochemistry*, **187**(2): 220-7. Olson, J.D., Panfili, P.R., Armenta, R., Femmel, M.B., Merrick, H., Gumperz, J., Goltz, M., and Zuk, R.F. (1990). A silic n sensor-based filtration immunoassay using biotin-mediated capture. *Journal of Immunological Methods*, **134**(1): 71-9. Olson, J.D., Panfili, P.R., Zuk, R.F., and Sheldon, E.L. (1991). Quantitation of DNA hybridization in a silicon sensor-based system: application to PCR. *Molecular & Cellular Probes*, **5**(5): 351-8.), incorporated by reference, and immunochemical reagents such as antibody-enzyme conjugates (Eberle, G., Barbin, A., Laib, R.J., Ciroussel, F., Thomale, J., Bartsch, H., and Rajewsky, M.F. (1989). 1,N6-etheno-2'-deoxyadenosine and 3,N4-etheno-2'-deoxycytidine detected by monoclonal antibodies in lung and liver DNA of rats exposed

to vinyl chloride. *Carcinogenesis*, **10**(1): 209-12. Foiles, P.G., Miglietta, L.M., Nishikawa, A., Kusmierek, J.T., Singer, B., and Chung, F.L. (1993). Development of monoclonal antibodies specific for 1,N2-ethenodeoxyguanosine and N2,3-ethenodeoxyguanosine and their use for quantitation of adducts in G12 cells exposed to chloroacetaldehyde. *Carcinogenesis*, **14**(1): 113-6. Palecek, E., and Hung, M.A. (1983). Determination of nanogram quantities of osmium-labeled nucleic acids by stripping (inverse) voltammetry. *Analytical Biochemistry*, **132**(2): 236-42.), incorporated by reference.

In this Step 4a detection, the etheno derivatives (primarily the ethenoadenine residues) within the loops are measured with the fluorimeter of the apparatus: excitation at 310 nm, and emission at 410 nm. The degree of fluorescence and sensititivity of the fluorimeter is calibrated with a quinine sulfate standard (10^{-5} - 10^{-7} M in 0.1 N H_2SO_4). The amount of direct etheno fluorescence is increased by a factor of 2 by completely digesting the samples with DNaseI and phosphodiesterase, when a gel overlay is used to prevent diffusion of the signals and disruption of the two-dimensional array of markers. The number of heteroduplexes is determined by the unique fluorescence of the adduct that was initially linked to the Left primers. Rhodamine, fluorescein or isothiocyanine derivatives can all be used to obtain intense fluorescent signals that can be separately measured from the fluorescence of the etheno adducts. Standard programs quantitate the two different signals by analyzing two or more regions of the emission and/or excitation spectra. Alternative detection methods for the etheno- derivatives include the use of specific monoclonal

antibodies (Eberle, G., Barbin, A., Laib, R.J., Ciroussel, F., Thomale, J., Bartsch, H., and Rajewsky, M.F. (1989). 1,N6-etheno-2'-deoxyadenosine and 3,N4-etheno-2'-deoxycytidine detected by monoclonal antibodies in lung and liver DNA of rats exposed to vinyl chloride. *Carcinogenesis*, 10(1): 209-12. Foiles, P.G., Miglietta, L.M., Nishikawa, A., Kusmierek, J.T., Singer, B., and Chung, F.L. (1993). Development of monoclonal antibodies specific for 1,N2-ethenodeoxyguanosine and N2,3-ethenodeoxyguanosine and their use for quantitation of adducts in G12 cells exposed to chloroacetaldehyde. *Carcinogenesis*, 14(1): 113-6.), incorporated by reference, conjugated to a chemiluminescent (including horseradish peroxidase or betagalactosidase) or electrochemical (urease and silicon-detector system) detection method.

In this Step 4a detection, residues within a mismatch loop will display differing degrees of reactivity to the modifying reagents as well as interactions (including fluorescence quenching and energy transfer) between closely spaced ethenoderivatives. (Which is why the fluorescence of an etheno derivative in a polynucleotide is approximately half that of the free ethenonucleotide.) Thus, systematic labeling is used to calibrate the fluorescent signal for each size of mismatch loop, thereby compensating for the nonlinearity of the fluorescent signal with respect to the loop size.

In Step 4a*, an alternative embodiment of the heteroduplex loop detection is accomplished by incorporating labeled nucleotides during the Step 2a PCR synthesis, and then in St p 31* digesting them out of the single-stranded loops of the heteroduplex. Incorporating labeled nucleotides

(e.g., fluorescently or radioactively, using appropriate triphosphate deoxynucleotide precursors) has greater signal strength than, and is therefore preferrable to, direct measurement of the liberated nucleotides by optical density. The quantity of detectable freed label corresponds to the loop size. This is done using a single-strand specific nuclease, such as S1 nuclease from Aspergillus orcyze (Dodgson, J.B., and Wells, R.D. (1977). Action of single-strand specific nucleases on model DNA heteroduplexes of defined size and sequence. *Biochemistry*, **16**(11): 2374-9. Gite, S., and Shankar, V. (1992). Characterization of S1 nuclease. Involvement of carboxylate groups in metal binding. *European Journal of Biochemistry*, **210**(2): 437-41. Shenk, T.E., Rhodes, C., Rigby, P.W., and Berg, P. (1975). Biochemical method for mapping mutational alterations in DNA with S1 nuclease: the location of deletions and temperature-sensitive mutations in simian virus 40. *Proceedings of the National Academy of Sciences of the United States of America*, **72**(3): 989-93. Wiegand, R.C., Godson, G.N., and Radding, C.M. (1975). Specificity of the S1 nuclease from Aspergillus oryzae. *Journal of Biological Chemistry*, **250**(22): 8848-55.), incorporated by reference, native micrococcal nuclease (Chambers, S.A., and Rill, R.L. (1984). Enrichment of transcribed and newly replicated DNA in soluble chromatin released from nuclei by mild micrococcal nuclease digestion. *Biochimica Et Biophysica Acta*, **782**(2): 202-9. Galcheva, G.Z., Davidov, V., and Dessev, G. (1985). Formation of single-stranded regions in the course of digestion of DNA with DNAase II and micrococcal nuclease. *Archives of Biochemistry & Biophysics*, **240**(1): 464-9.), incorporated by reference, or modified micrococcal nuclease (Corey, D.R., Pei, D., and Schultz, P.G. (1989). Generation of a catalytic sequence-specific hybrid DNase. *Biochemistry*, **28**(21): 8277-

86. Pei, D., Corey, D.R., and Schultz, P.G. (1990). Site-specific cleavage of duplex DNA by a semisynthetic nuclease via triple-helix formation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24): 9858-62.), incorporated by reference. When an apparatus is used that permits comingling of contents from different chambers, the spatial separation of the released nucleotides is maintained by performing the nuclease reaction in a gel overlay of the polystyrene plate. The gel prevents diffusion of the released nucleotides. (Diffusion is not an issue with direct detection of chemically modified nucleotides.) Alternatively, the polystyrene plate is placed into a plastic manifold, recreating 384 separate chambers.

In an alternative embodiment of the Step 4 detection, chemical modification is combined with specific nuclease treatment. S1 or micrococcal nuclease can be used to enhance the fluorescence of the etheno-derivatized adenosines generated by the chloracetaldehyde reaction. This provides two sets of measures of the same residues, thus increasing accuracy and sensitivity. The nuclease treatment can be used alone to liberate nucleotides from the loop. These free nucleotides are then separated from the retained double-stranded DNA of the heteroduplexes and quantitated. The spatial orientation of the reactions must be preserved as the nucleotides are released. This is done by performing the nuclease reaction in a gel, such as polyacrylamide that is on a solid backing (available from FMC Corporation), or by fitting a manifold over the streptavidin plate to contain the solutions with the nuclease and free nucleotides. To use the polyacrylamide gel plate, one takes a 0.1 - 0.5 mm polyacrylamide gel (ranging from 4-15%) bound to a plastic backing. The gel is slightly dehydrated with minimal surface

moisture. The nuclease solution is applied to the surface of the gel (the amount of S1 or micrococcal nuclease must be titrated for the enzyme lot) and the gel is placed over the surface of the streptavidin plate to which the heteroduplexes are bound. After incubating for 10-45 minutes (preferably 15 minutes) at room temperature to 37°C (preferably at 37°C), the gel layer is removed and the nucleotides embedded within the gel are quantitated by fluorescence, two-dimensional radioactivity counting, autoradiography, or immunochemical assays.

An alternative detection mechanism is described in Steps 4b, 4c, and 4d. The nucleotides within the heteroduplex loops are detected by distinguishing these nucleotides fr m those that are contained within the double-stranded portions of the DNA strands. In the preferred embodiment, the chemical modification agent chloracetaldehyde that selectively reacts with the exposed nucleotides within the loops is employed to specifically modify the C and A nucleotides within th heteroduplex loops. This reagent is preferrable to other chemical modification agents such as hydroxylamine, bisulfite, and osmium tetroxide because of its ease of use, and the fact that the derivatized nucleotides are fluorescent, while the chemical reagent and the unmodified nucleotides are not fluorescent. These other chemical methods represent alternative embodiments, with reagent conditions and detection methods adjusted accordingly as described by established techniques (Cotton, R.G. (1993). Current methods of mutation detection. *Mutation Research*, **285**(1): 125-44. Ganguly, A., and Prockop, D.J. (1990). Detection of single-base mutations by reaction of DNA heteroduplexes with a water-soluble carbodiimide followed by primer extension: application to products from the polymerase

chain reaction. *Nucleic Acids Research*, 18(13): 3933-9. Glikin, G.C., Vojtiskova, M., Rena, D.L., and Palecek, E. (1984). Osmium tetroxide: a new probe for site-specific distortions in supercoiled DNAs. *Nucleic Acids Research*, 12(3): 1725-35. Hayatsu, H. (1976). Reaction of cytidine with semicarbazide in the presence of bisulfite. A rapid modification specific for single-stranded polynucleotide. *Biochemistry*, 15(12): 2677-82. Jelen, F., Karlovsky, P., Makaturova, E., Pecinka, P., and Palecek, E. (1991). Osmium tetroxide reactivity of DNA bases in nucleotide sequencing and probing of DNA structure. *General Physiology & Biophysics*, 10(5): 461-73. Lilley, D.M. (1983). Structural perturbation in supercoiled DNA: hypersensitivity t modification by a single-strand-selective chemical reagent conferred by inverted repeat sequences. *Nucleic Acids Research*, 11(10): 3097-112. Smooker, P.M., and Cotton, R.G. (1993). The use of chemical reagents in the detection of DNA mutations. *Mutation Research*, 288(1): 65-77. Tindall, K.R., and Whitaker, R.A. (1991). Rapid localization of point mutations in PCR products by chemical (HOT) modification. *Environmental & Molecular Mutagenesis*, 18(4): 231-8.), incorporated by reference. In an alternative embodiment, a detection amplification method such as immunodetection of the adducts using a urease-conjugate and a silicon-based detection of a pH shift, contacts the polystyrene surface with an electronic silicon detector and a urea-containing gel interface using existing methods (Briggs, J., Kung, V.T., Gomez, B., Kasper, K.C., Nagainis, P.A., Masino, R.S., Rice, L.S., Zuk, R.F., and Ghazarossian, V.E. (1990). Sub-femtomole quantitation of proteins with Threshold, for the biopharmaceutical industry. *Biotechniques*, 9(5): 598-606. Kung, V.T., Panfili, P.R., Sheldon, E.L., King, R.S., Nagainis, P.A., Gomez, B.J., Ross, D.A., Briggs, J., and Zuk,

R.F. (1990). Picogram quantitation of total DNA using DNA-binding proteins in a silicon sensor-based system. *Analytical Biochemistry*, **187**(2): 220-7. Olson, J.D., Panfili, P.R., Armenta, R., Femmel, M.B., Merrick, H., Gumperz, J., Goltz, M., and Zuk, R.F. (1990). A silicon sensor-based filtration immunoassay using biotin-mediated capture. *Journal of Immunological Methods*, **134**(1): 71-9. Olson, J.D., Panfili, P.R., Zuk, R.F., and Sheldon, E.L. (1991). Quantitation of DNA hybridization in a silicon sensor-based system: application to PCR. *Molecular & Cellular Probes*, **5**(5): 351-8.), incorporated by reference.

In Step 5, the genotypes are determined for every STR. The two signals for each locus represent the sum and difference between the alleles. When compared with predetermined calibration tables, this representation becomes quantitative. One allele is computed by adding the sum and difference values and then dividing by two, and the second allele is computed by subtracting the sum and difference values and then dividing by two. This genotype determination is done for every locus.

While the foregoing method has been described for the measurement of loop mismatches as a technique for distinguishing the alleles STRs, the same approach is applicable to detecting specific gene alleles and mutations. For mutation detection, chemical modification by CAA as well as by other reagents at the site of the basepair mismatch creates a detectable signal. The use of a bound oligonucleotide to create a solid-state detection of specific alleles has been described (Giorda, R., Lampasona, V., Kocova, M., and Trucco, M. 1993. Non-Radioisotopic Typing of Human Leukocyte Antigen Class II Genes on Microplates.

BioTechniques, 15(5): 918-925. Lemna, W.K., Feldman, G.L., Kerem, B.-S., Fernbach, S.D., Zevkovich, E.P., O'Brien, W.E., Riordan, J.R., Collins, F.S., Tsui, L.-C., and Beaudet, A.L. 1990. Mutation analysis for heterozygote detection and the prenatal diagnosis of cystic fibrosis. *N. E. J. Med.*, 322: 291-296.), incorporated by reference, and the use of chemical reagents to modify the sites of basepair mismatch is also well-described (Cotton, R.G. (1993). Current methods of mutation detection. *Mutation Research*, 285(1): 125-44.), incorporated by reference. The invention described herein combines chemical modification techniques with solid-state detection in a novel manner different from any existing gel electrophoresis method.

From the resulting dense genotyping data, the descent of chromosomal segments within families and populations can be traced. This is because the number of recombinations is small compared with the linear sampling density of the chromosomes. Hence, agreement of alleles at many consecutive closely-spaced markers having a high polymorphism information content (PIC) value (Botstein, D., White, R.L., Skolnick, M.H., and Davies, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.*, 32: 314-31.), incorporated by reference, serves as a signature (with extremely high probability) in the population for a unique linear segment of chromosome. In fact, with sufficiently dense spacing (as described here), loci having much less informative PIC values can be used.

Phenotypic data is gathered on the individuals, animals, or plants which are genotyped. For humans, this includes the basic medical examination: history, physical, and laboratory data. Additional phenotypic markers for various genetic

diseases (e.g., creatine kinase for Duchenne muscular dystrophy) can also be collected. Environmental risks and exposures are also recorded.

These genes associated with phenotypic traits are then localized on the genome. This analysis can be done by linkage (Ott, J. 1991. *Analysis of Human Genetic Linkage, Revised Edition*. Baltimore, Maryland: The Johns Hopkins University Press. Feingold, E., Brown, P.O., and Siegmund, D. 1993. Gaussian Models for Genetic Linkage Analysis Using Complete High-Resolution Maps of Identity by Descent. *Am. J. Hum. Genet.*, 53: 234-252.), incorporated by reference, affected pedigree member (Weeks, D.E., and Lange, K. 1988. The affected pedigree member method of linkage analysis. *Am. J. Hum. Genet.*, 42: 315-326. Weeks, D.E., and Lange, K. 1992. A multilocus extension of the affected-pedigree-member method of linkage analysis. *Am. J. Hum. Genet.*, 50: 859-868.), incorporated by reference, affected relative pairs (Risch, N. 1990. Linkage strategies for genetically complex traits. (in three parts). *Am. J. Hum. Genet.*, 46: 222-253.), incorporat d by reference, inclusion/exclusion, (Perlin, M.W., and Chakravarti, A. 1993. Efficient Construction of High-Resolution Physical Maps from Yeast Artificial Chromosomes using Radiation Hybrids: Inner Product Mapping. *Genomics*, 18: 283-289.), incorporated by reference, association, homozygosity mapping (Ben Hamida, C., Doerlinger, N., Belal, S., Linder, C., and Reutenauer, L. 1993. Localization f Friedrich ataxia phenotype with selective vitamin E deficiency to chromosome 8q by homozygosity mapping. *Nature Genetics*, 5: 195-200. Pollak, M.R., Chou, Y.-H.W., Cerda, J.J., Steinmann, B., LaDu, B.N., Seidman, J.G., and Seidman, C.E. 1993. Homozygosity mapping of the gene for alkaptonuria to chromosome 3q2. *Nature Genetics*, 5(201-4).), incorporated

by reference, linkage disequilibrium, and other genetic localization techniques (Emery, A.E.H. 1986. *Methodology in Medical Genetics: an introduction to statistical methods, Second Edition*. Edinburgh: Churchill Livingstone. Vogel, F., and Motulsky, A.G. 1986. *Human genetics: Problems and Approaches, Second Edition*. Berlin: Springer-Verlag.), incorporated by reference. The result is one or more (with polygenic disease) peaks appearing at specific locations on the chromosome that both suggest specific gene regions, as well provide a signature pattern for phenotypic risk. With dense STS sampling along the genome (i.e., x-axis), and large numbers of individuals tested at these STSs, with each STS's allele given a combined score (i.e., on a y-axis), the conventional limitations of statistical linkage analysis are overcome, and the process becomes akin to a signal processing of genetic data in order to separate delta functions (i.e., the causative genes) from the background noise. That is, in addition to conventional linkage analysis, a method based on superimposing genetic information from many related individuals as one dimensional signals (along a genome) will accurately identify recurring genome locations by where th peaks occur. This method is described in figure 12. Importantly, this methodology will work well with complex multigenic multifactorial diseases (Lander, E.S., and Botstein, D. 1986. Mapping Complex Genetic Traits in Humans: New Methods Using a Complete RFLP Linkage Map. In *Cold Spring Harbor Symposia on Quantitative Biology*, 49-62. vol. LI, Cold Spring Harbor, Cold Spring Harbor Laboratory.), incorporated by reference, and not just single gene Mendelian inherited diseases. These complex diseases include all the most comm n diseases, such as cancer, heart disease, vascular disease, diabetes, glaucoma, and lung disease (King, R.A., Rotter, J.I., and Motulsky, A.G., ed. 1992. *The Genetic Basis of*

Common Diseases. New York, NY: Oxford University Press.), incorporated by reference.

Risks of trait inheritance or disease can then b determined by probabilistic (e.g., Bayesian) techniques (Young, I.D. 1991. *Introduction to Risk Calculation in Genetic Counselling*. Oxford: Oxford University Press.), incorporated by reference, that correlate the available genotypic and phenotypic data and environmental factors with chance of disease occurrence. In particular, the signatures of causative gene locations deduced from the population can be applied to each individual to ascertain risk. For animal and plant studies, one or more genetic loci can be associated with specific (desirable or undesirable) traits such as milk production or disease resistance. This information can be used for selective breeding.

Once the risks have been computed for an individual (in the context of his or her family) for all known disease entities, they can be sorted in descending order of likelihood and severity. The entities appearing at the top of this list are precisely those diseases that this individual has the greatest risk of developing. By moderating the environmental factors of these entities, including diagnostic, therapeutic, and preventative measures, the risks of these diseases can be reduced. This enables true cost-effective implementation of preventive health care: full customization to the genomic composition of each patient.

The techniques of genotyping and phenotypic correlation can be similarly applied to the task of disease gene identification. Exploiting dense genotypic data is particularly advantageous over existing techniques in

localizing the genes of complex multigenic diseases. Onc genes have been localized on the genetic map, use of an integrated genetic/physical genome map allows the positional cloning (Kerem, B.-S., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M., and Tsui, L.-C. 1989. Identification of the cystic fibrosis gene: genetic analysis. *Science*, 245: 1073-1080. Riordan, J.R., Rommens, J.M., Kerem, B.-S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M.L., Iannuzzi, M.C., Collins, F.S., and Tsui, L.-C. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, 245: 1066-1073.), incorporated by reference, of the causative genomic materials. As more genes are mapped, the task increasingly becomes the association between known genes with specific traits and disease, rather than the isolation of new genes.

The storage and safeguarding of this genetic information requires large, secure memory devices. These can restrict access to just those persons given authority by the individual. In one embodiment, individuals are given CD-ROMs containing the genetic information of themselves and their relatives, with access restricted by encryption and passwords, so that each individual can only directly grant access to information about themselves. Alternatively, a centralized data service can provide this secure information.

With the large amount of genotypic, phenotypic, and risk assessment data obtained, the results of the customized risk analysis must be presented in a coherent fashion to the patient. This is done with the assistance of genetic counselors (Emery, A.E.H., and Rimoin, D.L., ed. 1983.

Principles and practice of medical genetics. Edinburgh: Churchill Livingstone.), incorporated by reference, or clinical geneticists, or with a computer-based system that replicates this expertise. What is essential is to keep the bulk of the megabytes of information and low risk diseases in the background, and only bring to the patient's attention the most relevant risk and prevention information.

D. Method for Genotyping STRs using Loop Mismatch

The STR loop mismatch method employs heteroduplex hybridizations to directly measure the STR allele repeat number n. Consider the two alleles at a given STR locus as the complementary strands in a heteroduplex DNA molecule. Suppose that one strand S contains s STR repeat units and its mismatched complementary strand T' contains t STR repeat units. (Notation: U' denotes the complementary strand of sequence U.) Each STR repeat unit is comprised of k nucleotides. Assume that the left and right flanking regions are identical (i.e., perfectly complementary). When the hybridization product ST' is formed, if s=t (i.e., identical STR alleles), then there is a perfect match of the duplex DNA. If, however, $s \neq t$ (i.e., different STR alleles), then a heteroduplex is formed that has a loop of single-stranded nucleic acid (SS-DNA).

D.1. Method for Genotyping STRs using Loop Modification

Referring to figure 4, for s>t, the loop structure seen in subfigure 4A is formed. Here, subsequence L 402 is the left flanking region (with subsequence L' 404 complementary) and subsequence R 406 is the right flanking region (with subsequence R' 408 complementary). Crucially, the s-t extra

STR units form a single-stranded loop 410 of size (s-t)*k bases. Energetically, only one such loop is expected (Ninio, J. 1979. *Biochimie*, 61: 1133. Salser 1977. *Cold Spring Harbor Symp. Quant. Biol.*, 42: 985.), incorporated by reference; however, multiple loops would in no way change the results.

For s<t, the complementary structure shown in subfigure 4B is formed. Here, the single-stranded loop 412 size (t-s)*k is on the complementary strand.

The key idea is this: by detecting the size of the single-stranded loop 410 or 412, the value s-t (or t-s) can be determined. By comparing two unknown alleles with a known standard, and by also comparing the two alleles with respect to each other, these loop size measurements will precisely determine the two alleles, i.e., the genotype at the STR locus.

The signal strength from a loop of single-stranded DNA is proportional to the number of unmatched nucleotides in the heteroduplex ST'. This signal is measured by means of a first label (*) that corresponds to the number of unmatched nucleotides in the loop of ST'. This label is measured by means of a physical detection that preferentially detects specific nucleotides in single-stranded DNA.

In the most preferred "chemical modification" embodiment, the nucleotides in the S strand of the heteroduplex molecule are chemically modified after the PCR synthesis. The modification to these nucleotides renders them detectable (e.g., by fluorescence). The measured fluorescence of these modified S nucleotides is proportional to the size of the loop mismatch s-t.

In another preferred "synthesis/digestion" embodiment, the nucleotides in the S strand of the heteroduplex molecule are labeled (radiolabeled, or other detectable means) and then incorporated during the PCR synthesis. Subsequent digestion with an S1-like endonuclease separates the mismatched (and labeled) S nucleotides from the heteroduplex. The measured signal of these released S nucleotides is proportional to the size s-t of the loop mismatch prior to enzymatic digestion.

Means of physical detecting a quantitative signal for determining the loop size include: radioactivity, fluorescence, optical density, ionic concentration, electromagnetic conductivity or susceptibility, electrochemical coupling, or other detection assays (all referred to previously in this description).

The loop size is determined by the ratio of the (1) measured single-stranded loop signal strength to the (2) measured number of strands having a loop. Therefore, in addition to detecting loop size, accurate quantitation also requires determining the number of heteroduplex strands with measurable loops. This is done using an independent second label (#) on the S strands of the heteroduplex molecules. This label is comprised of a detectable molecule attached to the PCR primer of the S strand; subsequent measurement of this molecule quantifies the number of strands in heteroduplexes.

Although this loop mismatch method applies to all VNTRs of the form LW_nR, the following discussion assumes throughout an STR with W="CA". This is done solely to clarify the presentation, since the loop mismatch approach will work with any STR or VNTR locus, and on any linear nucleic acid (i.e., DNA, RNA, and hybrid polymers).

To further clarify the presentation, the loop with label (*) is indicated by A*s, which in the most preferred embodiment represent adenosine nucleotides on the single-stranded loop that are chemically modified by chloracetaldehyde into a detectable state. The presentation is written to be compatible with another preferred embodiment, wherein the A*s represent labeled (e.g., radiolabeled) nucleotides that are incorporated during PCR synthesis, and are then detected following endonucleas digestion.

To determine a single allele (e.g., homozygous or hemizygous locus), the experiment consists of performing a PCR amplification of an unknown CA-repeat locus source S of the form $L(CA)_nR$, and hybridizing it to a known complementary oligonucleotide target T' of the form $[L(CA)_nR]'$ in order to indu ce mismatch and quantitatively measure the loop.

Referring to figure 5, in Step 1 a CA-repeat locus molecule is selected for analysis, and is defined by its unique left and right oligonucleotide primers. The primers are synthesized with appropriate labeling and linking modifications (Haralambidis, J., Duncan, L., Angus, K., and Tregear, G.W. (1990). The synthesis of polyamide-oligonucleotide conjugate molecules. *Nucleic Acids Research*,

18(3): 493-9. Nelson, P.S., Kent, M., and Muthini, S. (1992). Oligonucleotide labeling methods. 3. Direct labeling of oligonucleotides employing a novel, non-nucleosidic, 2-aminobutyl-1,3-propanediol backbone. *Nucleic Acids Research*, 20(23): 6253-9. Roget, A., Bazin, H., and Teoule, R. (1989). Synthesis and use of labelled nucleoside phosphoramidite building blocks bearing a reporter group: biotinyl, dinitrophenyl, pyrenyl and dansyl. *Nucleic Acids Research*, 17(19): 7643-51. Schubert, F., Cech, D., Reinhardt, R., and Wiesner, P. (1992). Fluorescent labelling of sequencing primers for automated oligonucleotide synthesis. *Dna Sequence*, 2(5): 273-9. Theisen, P., McCollum, C., and Andrus, A. (1992). Fluorescent dye phosphoramidite labelling of oligonucleotides. *Nucleic Acids Symposium Series*, 1992(27): 99-100.), incorporated by reference.

In Step 2a, the target DNA TT' is constructed from a standard of known CA-repeat length t in a separate PCR experiment. The allele size t is chosen sufficiently small, say between 0 and 10, so that s>t is always guaranteed. Standard PCR amplification of genomically-derived or cloned DNA for 20-40 cycles is done using unlabeled primers and nucleotides, with a linker such as biotin on the right primer.

In Step 2b the source DNA SS' is constructed from sampl genomic DNA via a PCR experiment. The CA-repeat locus molecule is defined by its unique left and right primers. A standard PCR amplification of genomically derived DNA is done for 20-40 cycles using labeled (#) left primer in the presence of A* labeled nucleotides.

In Step 3a the SS' and TT' duplex molecules are denatured to form single stranded DNAs. When renatured in solution, the hybridization pairs

(S+T)(S+T)'

recombine to form

SS', ST', TS', and TT'.

The T strands of the TT' duplex are not detectable (since their loops match), and can be factored out of the analysis. For example, the T strands can be removed by attaching the TT' duplex to solid support via the linker of T', and then denaturing T from T', and washing to remove T, thus purifying T'. Alternatively, T can remain as a nondetectable competitive contaminant. Further, using an excess of SS' relative to TT' favors the production of ST' heteroduplexes. Therefore, the focus is on the hybridization pairs

(S)(S+T)'

which recombine to form

SS' + ST'.

The SS' contains no single-stranded loops, hence is not detectable. Further, since only the T' molecule has the linker for solid support, attaching the T' to a surface (e.g., the biotin of T' to a streptavidin-coated surface) and washing removes the SS' product. This leaves only

ST'

as a detectable (and useful) product.

Referring to subfigure 5A, the heteroduplex molecule is comprised of an upper strand 502 and a complementary lower strand 504. With s>t, the hybridization product is as shown in subfigure 5A.

(S) The upper source strand 502 is produced by a first PCR amplification of sample genomic DNA.

(S1) The single-stranded DNA loop 506 contains *-detectable A nucleotides. (Following chemical modification or by incorporation/digestion, the *-detectable A's are used to measure loop size via label *.)

(S2) A second label (#) 508 on the upper strand is for strand quantification, and is attached to the left PCR primer.

(T') The complementary lower target strand 504 is produced by a different PCR amplification of a known STR locus, or by direct synthesis. This lower strand has a linker 510 such as biotin attached to its 5' (right) end.

With s>t, the hybridization of strands S and T' are perfectly matched everywhere but in the CA-repeat region. This mismatch produces a loop of size 2(s-t) containing precisely (s-t) A*'s. In the Step 3b chemical modification embodiment, the exposed A*s on the single-stranded DNA loop are chemically modified by chloracetaldehyde, as shown in

subfigure 5B; in Step 4a, detecting the fluorescence from the first label (*) on the A* 512 measures the magnitude of s-t.

In the alternative Step 3c synthesis/digestion embodiment, the exposed A*s on the single-stranded DNA loop are digested from the heteroduplex into free A* 514 using an endonuclease, as shown in subfigure 5C; in Step 4b, the radioactive A* is then detected using a scintillation counter, thereby measuring the magnitude of s-t.

Only the upper S strands 516 have the second label (#) 518, so detection in this fluorescent molecule's wavelength in Step 4a or 4b measures the number of strands.

The allele is determined in Step 5. Calibrations done prior to the experiment ensure that these measurements provide precise quantitation. Since

label 1 (*) => (number of SS-DNA strands) * (s-t), and
label 2 (#) => (number of SS-DNA strands),

taking the calibrated ratio of label1 (*) to label2 (#) gives a measure of s-t. When only one allele s is measured (as in a hemizygotic or homozygotic locus, or with separated chromosomes), this determines the value s-t. Since s>t, the allele s is then determined by adding the known value t to the measured value (s-t).

The determination of the allele sum s1+s2 is described next. For general genotyping, the heterozygotic case must be handled. Suppose that a CA-repeat locus is heterozygotic, comprised of two alleles having CA-repeat numbers s1 and s2 corresponding to their respective DNA strands S1 and S2.

Referring to figure 5, performing the Steps 1 and 2 of the two PCR experiments and the Step 3 hybridization with strand T', the two products

(S1+S2)T' ,

or

S1,T'; and S2,T'

are formed. These two species are present in equal concentrations.

Step 4 measures the sum $s=[(s1-t)+(s2-t)]/2$. Following calibration, Step 5 adds the known value t to s, forming th average (s1+s2)/2 of the alleles. Multiplying this average by 2 determines the allele sum s1+s2.

Determining the allele difference s2-s1. This experiment consists of performing a PCR amplification of an unknown CA-repeat locus with the (zero, one, or) two sources S1 and S2 of the form $L(CA)_{s1}R$ and $L(CA)_{s2}R$, and hybridizing them against each other's complementary strands. This induces a loop mismatch proportional to $|s2-s1|$, which is then quantitatively measured.

Referring to figure 6, in Step 1 a CA-repeat locus molecule is selected for analysis, and is defined by its unique left and right oligonucleotide primers. The primers are located far enough away from the CA-repeat region to assure a sufficiently long linear stretch of DNA in the homoduplex; this is done make the effect of different loop sizes on the free energy neglible. The rationale is that

the flanking regions and the complementary CA/GT repeat regions have a total free energy that is proportational to the number of matching nucleotides, whereas the single-stranded DNA loop of heteroduplex has a free energy that grows as the logarithm of the loop size (Ninio, J. 1979. *Biochimie*, 61: 1133. Salser 1977. *Cold Spring Harbor Symp. Quant. Biol.*, 42: 985.), incorporated by reference. Thus, relative to the large region of matched double-stranded DNA, the free energy changes (and binding affinities) introduced by differing loop sizes is small.

In optional Step 2a, target DNA TT' is constructed from a standard of known CA-repeat length t in a separate PCR experiment. The allele size t is chosen sufficiently small, say between 0 and 10, so that s>t is always guaranteed. Standard PCR amplification of genomically-derived or cloned DNA for 20-40 cycles is done using unlabeled primers and nucleotides. No labels or linkers are used.

In Step 2b, the two source alleles are constructed simultaneously in one PCR experiment: each allele serves as the hybridization target for the other. A standard PCR amplification of genomically derived DNA is done for 20-40 cycles using labeled (#) left primer, and a right primer with a linker such as biotin, in the presence of A* labeled nucleotides.

Step 3a forms the heteroduplexes. The S1,S1' and S2,S2' homoduplex molecules are denatured to form single stranded DNAs. When renatured in solution, the hybridization pairs

(S1+S2)(S1+S2)'

recombine to form the four products

S1,S1'; S1,S2'; S2,S1'; and S2,S2'.

All four species are present in roughly equal concentrations. This is because of the DNA energetics described in Step 1, which assures binding DNA affinities of approximately equal strength.

Referring to subfigure 6A, with s2>s1, the hybridization product is as shown. The heteroduplex molecule constructed after PCR amplifying the sample genomic DNA, and rehybridizing, is comprised of an upper strand 602 and a complementary lower strand 604.

(S) In the upper source strand 602:

(S1) The single-stranded DNA loop 606 contains *-detectable A nucleotides. (Following chemical modification or by incorporation/digestion, the *-detectable A's are used to measure loop size via label *.)

(S2) A second label (#) 608 on the upper strand is for strand quantification, and is attached to the left PCR primer.

(S') The lower strand, also has a linker 610 such as biotin attached to its 5' (right) end.

When s1=s2, S1 is the same molecule as s2, and the homoduplex S1,S1' is formed (the other three duplexes are equivalent). Since no mismatch occurs, there is no single-

stranded loop, and the detection measures zero signal, corresponding to the case s2-s1=0.

When $s1 \neq s2$, without loss of generality assume that $s1 < s2$. Consider the four hybridization cases:

(S1,S1') Homoduplex with no detectable signal.

(S2,S2') Homoduplex with no detectable signal.

(S1,S2') Since $s1 < s2$, the mismatch loop is on the S2' strand, and is unlabeled, producing no detectable signal.

(S2,S1') Since $s1 < s2$, the mismatch loop is on the S1 strand, and is labeled, producing a detectable signal.

(In another embodiment, the label is incorporated into both strands during the PCR by labeling the CA and/or the GT dN*'s. Hence, both the S1,S2' and S2,S1' strands have detectable single-stranded loops. Since both have the same $|s2-s1|$ loop size, there is a two- to four- fold increase in the desired measured signal.)

Incomplete hybridization results in single stranded lower DNAs S1' and S2' bound by biotin to the solid support. While there are no *-detectable As in the GT-repeat region of these lower strands, *-detectable A may be incorporated into the *flanking regions* during the PCR. In Step 3b these single-stranded segments are made nondetectable by DNA elimination and/or protection.

- *Elimination* can be done using a single-strand specific 3' to 5' exonuclease that removes SS-DNA

but not internal loops, such as E. coli exonuclease VII.

Protection is effected by generating nonlabeled upper T strands in Step 2a to form the double-stranded products

T,S1' and T,S2'.

The same short T strand with known allele size t (i.e., t<s1, and t<s2) used in figure 5 would work here in figure 6 as well. (Since t is smaller than both s1 and s2, the mismatch loops would be formed in the unlabeled GT-repeat region of the lower strands, hence would be undetectable.) Using just the left and right flanking regions L and R would also block the single-stranded flanking DNA, and have more favorable binding kinetics in that they would tend to not displace hybridized S1 and S2 strands.

These techniques can be combined for a more complete hybridization.

The hybridization of strands S2 and S1' is perfectly matched everywhere but in the CA-repeat region. This mismatch produces a loop of size 2*(s2-s1) containing precisely (s2-s1) A*'s. In Step 3c's chemical modification embodiment, the exposed A*s on the single-stranded DNA loop are chemically modified by chloracetaldehyde; in Step 4a detecting the fluorescence from the first label (*) on A*s measures the magnitude of s-t.

In Step 3d's alternative synthesis/digestion embodiment, the exposed A*s on the single-stranded DNA loop are digested

from the heteroduplex into free A* using an endonuclease; in Step 4b the radioactive A* is then detected using a scintillation counter, thereby measuring the magnitude of s-t.

All the strands (S1 and S2) are labeled with the fluorescent label (#), so Step 4a or 4b's detection in this fluorescent molecule's wavelength measures the total number of strands from all four hybrids.

In Step 5, the allele difference is determined. Calibrations done prior to the experiment assure that these measurements provide precise quantitation. Since

label 1 (*) => (strands/4) * (s2-s1), and
label 2 (#) => strands,

taking four times the calibrated ratio of label1 (*) to label2 (#) gives a measure of s2-s1.

The genotype is computed from loop mismatch data, referring to figure 7, by combining the sum (from the figure 5 protocol) and difference (from the figure 6 protocol) of the allele sizes; this determination exploits the elimination of PCR stutter artifact by pooling within each experiment, as described below. Thus, the single experiment of Step 1 accurately measures the allele sum (s1+s2), the single experiment of Step 2 accurately measures the allele difference |s2-s1|. Combining these in Step 3 determines the two alleles:

s1 = (sum - difference)/2
s2 = (sum + difference)/2

When fewer than two distinct alleles are present on the two chromosomes:
(0) zero alleles - both s1 and s2 are zero;
(1) one allele - s1 and s2 are equal (i.e., the difference is zero). The quantitation calibrated to other alleles shows whether one or two copies of the allele are present.

A detailed protocol is given for the loop mismatch method. The following steps referring to figure 8 are designed for measuring a single STR, rather than the multiple STRs assayed in figure 3. In Step 1 of figure 8, an STR locus is selected, and PCR primers are chosen to provide large flanking regions. In particular, this protocol is not optimized for compatibility with the apparatus of figure 1. The primers are synthesized derivatized to support the characterization experiments.

To genotype one STR, these modified primers are used:

- a first left primer L that is unmodified.
- a second left primer L# for the upper strand which has the flourescein label (#) at the 5' end,
- a first right primer R which contains no modifiers,
- a second right primer Rb containing one or more biotin residues at the 5' end or within the oligonucleotide.

Derivatizing the primer for binding to a surface entails incorporating a biotinylated nucleotide at the 5' end of the

synthetically made oligonucleotide. Additional biotinylated residues can be incorporated into this primer either at the time of biosynthesis or by secondary photo or chemical biotinylation. The preferred embodiment employs the direct addition of the 5' biotin by chemical synthesis; alternatively, additional biotin molecules may improve the heteroduplex isolation effiency.

In Step 2, three PCR amplifications are performed. Source DNA from a genome to be characterized, and target DNA of known minimal repeat length t from an individual (or prepared in advance by cloning a segment of genomic DNA in a plasmid or phage vector) are prepared for PCR. Three separate reactions are performed. These are identical, except for the following specific reaction mixtures:

PCR a: TT' sum PCR mixture for Step 2.a
target DNA, L, Rb, all dNTPs unlabeled

PCR b: SS' sum PCR mixture for Step 2.b
source DNA, L#, R, labeled a-^{32}P-dATP, other dNTPs unlabeled

PCR c: S2,S1' difference PCR mixture for Step 2.c
source DNA, L#, Rb, labeled a-^{32}P-dATP, other dNTPs unlabeled

The components of the PCR reaction are assembled so that each 0.2 or 0.5 ml tube contains the appropriate set of primers, followed by the standard PCR buffer containing Tris buffer, KCl, $MgCl_2$ and dNTP (the four triphosphat deoxynucleotides). The total size of each PCR reaction is 50 ul (though this can vary from 10-100 ul).

Each specific PCR reaction contains its specific reaction mixture, the PCR buffer (ie 10mM Tris pH8.0, 50 mM KCl, 2.5 mM magnesium chloride, albumin), and thermostable (e.g., Taq) polymerase. The PCR reaction is overlayed with a thin layer of Ampliwax that separates some of the components from each other so that the reaction begins when the temperature rises to a level that melts the wax and allows all of the components to mix. This is the "hot start" method of PCR which reduces nonspecific synthesis products. An initial heat denaturation of 93-95°C for 5 minutes is followed by the thermal cycles are performed 20-40 times. Each cycle consists of a 30 sec denaturation step at 95°C, 15-30 second annealing step at 50-65°C (typically 55°C) and an extension step at 73°C for 15-120 seconds (typically 45 seconds). When the three PCR reactions are completed, 0.5M EDTA is added to a final concentration of 10 mM. This inactivates the Taq polymerase.

In Step 3, the heteroduplex hybridizations and modifications are done. Reactions a and b are combined (summation experiment) in Step 3a, and reaction c (difference experiment) is kept separate in Step 3b. All the following operations are done independently for the two reactions (sum and difference). The samples are then heated to 95°C for 5 minutes and allowed to anneal at a temperature of 75°C to discourage primer-strand annealing. After 2-24 hours, the temperature is lowered to 4°C to solidify the Ampliwax and the exonuclease VII (Gibco, BRL), incorporated by reference, in the appropriate buffer is added to the surface. The buffer conditions for the PCR are compatible directly with those of exonuclease VII. The reactions are initiated by heating to 37°C and incubated for a time ranging from 1-120

minutes. The reactions are terminated by the addition of chloroform to the tubes.

The supernatants from the chloroform extractions contained hetero- and homoduplexes, digested single strands, primers and free nucleotides. The double-stranded DNA is then purified using a spin column/filter (such as Centricon filters from Amicon) to remove the small molecular weight material and concentrate the samples. The purified DNAs from experiment are then adsorbed to strepavidin paramagnetic beads (DYNAL 1993. Dynabeads biomagnetic separation system, Technical Handbook: Molecular Biology, Dynal International, Oslo, Norway.) to bind those double-stranded DNAs that contain the biotinylated right primer. The beads are washed several times with a neutral salt buffer to reduce nonspecific binding and not disrupt the double-stranded DNA.

In the preferred chemical modification embodiment of Step 3, the DNA bound to the strepavidin beads are equilibrated in 0.15 M Na Phosphate buffer pH = 6.5 (the pH can be varied from 4.5 to 6.5 and alternative buffers can be used) and then 2-chloracetaldehyde to a concentration 2.0%. The tubes are incubated at 37°C for 4 hours (longer or shorter times may be used). The reaction is terminated by washing with 0.01M Tris-HCl pH 7.0 and 1.0 M NaCl. The NaCl prevents dissociation of the heteroduplexes during the etheno-dehydration step. The samples heated in the final wash volume at 85°C for 1 hour (dehydrates the ethenoderivative).

In the alternative incorporation/digestion embodiment of Step 3, using a single-strand specific endonuclease such as S1 nuclease or micrococcal nuclease, the original PCR

products that have been treated with exonuclease and bound to the strepavidin beads are equilibrated in the endonuclease buffer and reacted for varying times.

In Steps 4a and 4b, the signals are detected. The fluorescence and radioactivity retained on the beads are measured. The amount of flourescein and ^{32}P can be independently determined. These values establish the number of double-stranded complexes, and total incorporation of ^{32}P cATP into the molecules. respectively.

In the preferred chemical modification embodiment, the fluorescence is measured by heating the samples to 95°C and eluting the DNA from the beads, taking the supernatents and measuring the fluorescence with a fluorimeter (excitation at 310 nm emission at 410 nm). The degree of fluorescence and sensititivity of the fluorimeter is calibrated with a quinine sulfate standard (10^{-5} - 10^{-7} M in 0.1 N H_2SO_4). The tubes can be counted again for the amount of retained flourescein and ^{32}P labels. The amount of radioactivity can be calibrated with known standards that account for tube geometry, sample volume and instrument counting efficiencies. Based upon the radioactivity and the fluorescence, the size of the loops can be established.

In the alternative incorporation/digestion embodiment, during the digestion, aliquots of the supernatants are removed and counted to determine the rate and extent of nuclease-dependent release of ^{32}P-labelled nucleotides. This establishes the optimal parameters for the endonuclease digestion and accurate quantitation of the nucleotides contained within the loops. This method can also be done

with the chloracetaldehyde-modified nucleotides to measure released fluorescence. A direct comparison of the two methods can be achieved with the same initial set of PCR reactions.

In Step 5, the genotype is determined. In Step 5a, the sum is computed from the Step 4a detection, and in Step 5b, the difference is computed from the Step 4b detection. The results are combined in Step 5c to determine the genotype of the STR, as described. This completes the protocol.

In another embodiment, DNA protection is done to minimize spurious signals from unhybridized single-stranded DNA, and exonucleases are not used. Referring to figure 8, in Step 3a, a 10:1 excess of the SS' amplified product relative to the TT' amplified product is preferably used. In Step 3b, when necessary, TT' (or fragments thereof) without labels or linkers is added to block unhybridized S1' strands.

In an alternative embodiment, the number of PCR reactions can be reduced by performing PCR reactions b and c above as a first reaction using a cleavable biotinylated right primer and modifying several steps. The PCR product can then be combined with the second target PCR reaction a to allow sequential measurement of the sum and difference experiments. This is accomplished by combining the two PCR reactions for the Source and Target DNA's in Step 2, preparing and isolating the heteroduplexes on the streptavidin beads in Step 3, and measuring the nucleotides within the loops by derivatization and fluorescence in Step 4. The initial measurements in Step 4 are then followed by the release of duplexes employing the immobilized Source strand by reduction of a disulfide linkage between the primer and the biotin. In Step 4, one measures the total number of

bound duplexes, the number of duplexes that are attributable to the biotinylated source, the total number of nucleotides contained within all loops and the number of nucleotides contained within the loops formed between the Target and Source DNA.

In an alternative embodiment, in Step 4, a more sensitive detection system for the chemical modification embodiment is an antibody-enzyme conjugate that recognizes the derivatized DNA (i.e., the etheno- derivatives created by chloracetaldehyde) and catalyzes a colorimetric reaction that can be measured in the supernatent. The simplest form of this assay would be to use a betagalactosidase-antibody conjugate that acts on a colorimetric substrate such as X-gal or Blue-gal (BRL/Gibco).

Eliminating PCR Stutter using Pooled Targets. When PCR is done on a CA-repeat locus, there is often a stutter pattern wherein smaller fragments are also generated in lesser amounts (Schwartz, L.S., Tarleton, J., Popovich, B., Seltzer, W.K., and Hoffman, E.P. 1992. Fluorescent Multiplex Linkage Analysis and Carrier Detection for Duchenne/Becker Muscular Dystrophy. *Am. J. Hum. Genet.*, 51: 721-729), incorporated by reference. With a locus $L(CA)_nR$, the fragments $L(CA)_{n-1}R$, $L(CA)_{n-2}R$, and so on are also generated in addition to the main PCR product $L(CA)_nR$. The distribution of the smaller fragments generally follows a decay pattern, with the amount of $L(CA)_mR$ less than $L(CA)_nR$, when $m<n$. This decay pattern is empirically observed to differ from one genetic locus to another, but remains stable across unrelated individuals for any given locus.

As described herein, the use of pooled targets in the preferred embodiment eliminates this artifact. Multiple sources hybridized against multiple targets, producing a quadratic number of heteroduplexes. The different CA-repeat sizes

s, s-1, s-2, ..., and
t, t-1, t-2, ...

are obtained for the DNA strands

S, S-1, S-2, ..., and
T', (T-1)', (T-2)', ... ,

which, when cross-hybridized, produce an entire table of products

(S-i) x (T-j).

The mismatch loop size of each hybrid (S-i)x(T-j) is (s-t-i+j).

The factors affecting the relative signal from each (S-i)(T-j) hybridization pair are:

(a) The product a_{ij} of the concentrations [S-i] and [T-j], which are determined by the stutter pattern. With equal amounts of S and T, and identical stutter patterns, the underlying concentration matrix is symmetric: $a_{ij} = a_{ji}$.

(b) The signal produced by the loop size of the mismatch, which is proportional to (or monotonic in) the length (s-t-i+j) of the mismatch.

(c) The differential amount of hybridization based on the energetics of DNA binding resulting from different loop sizes. As noted above, this is minor.

Combining the major factors (a) and (b), matrix symmetry results in the relative cancellation of off-diagonal terms. Each mismatch loop larger by d than the mean s-t is mirrored by a roughly equal concentration in its symmetric matrix entry of a mismatch loop smaller by d than the mean s-t. Thus, the total signal from the stuttered sources with the stuttered targets averages out to the mean value s-t.

This averaging by pooling with stuttered targets applies to all the aforementioned experiments.

- *Sum experiment*. The stuttered source (S1+S2) is hybridized with the complementary stuttered target T'. The stuttering is averaged away, and the desired signal strength (s1-t) + (s2-t) is measured.

- *Difference experiment*. The stuttered source (S1+S2) is hybridized with the complementary stuttered target (S1+S2)'. The four hybridization species occur. The mismatch loop length from the hybrid S2,S1' is formed from equally stuttered S2 and S1', so this measurement is correctly averaged.

Therefore, pooled experiments that use stuttered targets remove the stuttering from the signals.

When the measured signal is nonlinear in the loop size, factor (b) above would no longer be perfectly linear.

Nonetheless, the relationship between loop size and signal strength remains monotonic (and invertible). Calibration therefore removes stutter artifact.

STR Genotyping in a Combined Heteroduplex Experiment. The sum and difference experiments at a locus are done separately using separate PCRs: two for the sum, and one for the difference, as described above. The first PCR to construct TT' is preferably done prior to the introduction of sample genomic DNA, and can be incorporated (or "compiled") into the apparatus. Thus, with the introduction of sample DNA at "run time", the protocol of figure 8 employs two PCRs. The following describes how to reduce this to just one PCR experiment, thereby reducing operating time and space requirements. Digoxigenin is used as a linker (The Genius System User's Guide for Filter Hybridization, 1992. Boehringer Mannheim Corporation, Indianapolis, IN), incorporated by reference.

In an alternative embodiment, the sum and difference are obtained simultaneously in a combined PCR experiment. Referring to figure 9, in Step 1 an STR locus is selected and oligonucleotides prepared. In Step 2a, unlabeled duplex TT' of a known small repeat size t is constructed by PCR or direct synthesis. The right primer has a digoxigenin linker 1. In Step 2b, the homoduplexes S1,S1' and S2,S2' of an uncharacterized genomic DNA sample are amplified via PCR. The first label (*) is incorporated into the single-stranded loop, the left primer has the second label (#), and the right primer has a biotin linker 2. In Step 3, the duplexes are combined and denatured together at high temperature into their separate strands, yielding:

S2, S1, T, S2', S1', and T'.

Assume, as before, that $t<s1\leq s2$. Renaturing at a lower temperature forms the hybridization pairs between

(S2, S1, T) and (S2, S1, T)',

or the nine duplex DNA molecules arranged as the table:

S2,S2'	S2,S1'	S2,T'
S1,S2'	S1,S1'	S1,T'
T,S2'	T,S1'	T,T'

The detectabilities in Step 4 of the DNA hybridization pairs in this table are as follows:

- The upper right triangle submatrix hybrids provide all the detectable elements -

  (S2,S1') This gives the loop difference s2-s1.
  (S2,T') This gives one half of the loop sum s2-t.
  (S1,T') This gives the other half of the loop sum s1-t.

- The hybrid species (S2,S2'; S1,S1'; T,T') along the matrix diagonal are not detectable, since the duplex strands are identical in size, and no loop mismatch is formed.

- The lower left triangle submatrix hybrids are not detectable -

(S1,S2') By assumption, s1≤s2, so no loop mismatch is formed.

(T,S2') By assumption, t≤s2, so no loop mismatch is formed. However, this helps in blocking any unhybridized single-stranded DNA.

(T,S1') By assumption, t≤s1, so no loop mismatch is formed. However, this helps in blocking any unhybridized single-stranded DNA.

The T' (pre-made, digoxigenin linker 1) lower DNA strands from the S1' and S2' (locus-made, biotin linker 2) lower strands are spatially separated by using two different solid supports to specifically bind the digoxigenin and the biotin linkers in different measurable regions. Thus, the signals required for measuring the sum and the difference are detected in spatially separated experiments. In Step 5, the usual analysis (which exploits the expected PCR stuttering and the pooled targets) is used to compute the allele values.

In an alternative embodiment, a cleavable biotinylated linker is used on the right primer of T' that allows separate PCRs of a target and of genomic DNA, combines the samples into a single heteroduplex reaction, and then detects all nine of the hybridization products listed above. The following are measured: (a) the number of S1 and S2 strands bound, and (b) the number of nucleotides in the loops. Then, the S1,T' and S2,T' measurable heteroduplex species are liberated by reduction of the dissulfide linkage, followed by remeasuring the S2,S1' bound, and the number of nucleotides in the remaining loops.

A Scalable STR Genotyping Assay. The methods described refering to figure 8 enable practical construction of the apparatus in figure 1 and system manufactured device described in figure 3 in which multiple STR loci are genotyped simultaneously.

Of the five steps in figure 8, only Step (1) is specific for a given STR. The other four steps are largely independent of the given STR. Therefore, the apparatus in figure 1 is constructed to spatially encode multiple genetic loci on a surface, and places Step (1)'s specific STR oligonucleotides at each spatial location, prior to complete PCR processing. For the allele sum experiment Step (2a) in figure 8 deposits the pooled targets TT', and then Steps (2b-5) for the sample-dependent PCR processing, DNA hybridization, signal detection, and genotype determination are performed simultaneously over the surface. For the allele difference experiment Steps (2-5) for the sample-dependent PCR processing, DNA hybridization, signal detection, and genotype determination are performed simultaneously over the surface. In this way, the steps of figure 8 for single STR genotyping are related to the steps of figure 3 for multiple STR genotyping.

In an alternative embodiment, the spatiotemporal encoding of genetic loci is not restricted to a surface. Instead, the three dimensions of space and one dimension of time can be used to multiplex the STR-specific oligonucleotides and the PCR processing. For example, multiple reaction chambers in a three-dimensional arrangement would each contain STR-specific oligonucleotides over some time period. The PCR processing would be done in parallel in multiple chambers, until all required signals were obtain d.

This physical arrangement can customize the PCR conditions, if necessary, to each STR.

In a second example, commercially available 864-chamber plates can be physically arranged to achieve over 100,000 simultaneous characterizations. This is done by constructing a surface of four plates in a 2x2 array, which provides 3,456 chambers in a layer. Stacking thirty such layers provides 103,680 chambers. This three dimensional arrangement is quite compact, with no chamber further than two feet from any other chamber. For the amplification step, this three dimensional organization fits into a thermocycling PCR oven. The hybridization, detection, and other steps are multiplexed in time, enabling efficient use of the robotic device, detection device, and computer to achieve a throughput commensurate with the parallelization.

Double-Loop Detection for Improved Signal. In another embodiment, the signals from either the allele sum or allele difference experiments can be increased several-fold by detecting SS-DNA mismatch loops on *both* the upper and lower strands, rather than on just one strand. The PCR stutter can again be eliminated by using pooled targets.

The following description for determining a single allele refers to figures 5 and 10. The key change from the protocol referring to figure 5 is that nucleotides on both the upper and lower strands of S are made detectable. Step 1 of figure 5 selects the STR of interest, and prepares the oligonucleotides. The CA-repeat locus molecule is defined by its unique left and right primers, indicated in the figure by shading.

Step 2a of figure 5 amplifies the known homoduplexes TT'. Referring to subfigure 10A, the PCR primers 1002 (left) and 1004 (right) for the upper strand 1006 and the lower strand 1008 of the target TT' both contain linkers 1010 (e.g., biotin) for binding to solid support, but no (#) labels. The target TT' duplexes are constructed by standard PCR amplification of genomically derived DNA for 20-40 cycles using dNs without (*) labels.

Step 2b of figure 5 amplifies the unknown homoduplexes S1,S1' and S2,S2'. Referring to subfigure 10B, the first label (*) 1012 for loop quantitation is present on nucleotides (in equal proportions) in both strands S 1014 and S' 1016. The label (*) indicates detectability, whether by chemical modification or by incorporation/digestion. The second label (#) 1018 for strand quantitation is present on both the left 1020 and right 1022 PCR primers. The source DNA SS' is developed by standard PCR amplification of genomically derived DNA for 20-40 cycles using (*) labeled dA*, dC*, dG*, and dT*.

Following Step 3 of figure 5's denaturation and reannealing, the hybridization pairs formed are shown in the table of hybridization products of subfigure 10C in figure 10. These are:

(SS') Homoduplex is not detected, since no linker is present, and the loop size is zero.
(TT') Homoduplex is not detected, since the loop size is zero.
(ST') Each heteroduplex molecule has 2n loop size (*) labels, and one strand (#) label.

(TS') Symmetrically to ST', TS' is now also detected, with the same loop and strand labeling quantitation.

Thus, in detection Step 4 of figure 5's detection, 4n loop size (*) signals, and 2 strand (#) signals are measured per ST' molecule. In Step 5's allele determination, this four-fold increase in label (*) and two-fold increase in label (#) is accounted for.

The analysis of the Steps in figure 5 applies to the case of two alleles S1 and S2 for determining the allele sum. Two separate PCRs are done, as described: one for S1,S1' and S2,S2' labeled duplexes, and one for linker TT' targets. By ensuring that s>t, the denaturation/reannealing experiment constructs nine hybridization products. However, only those containing an T or T' linker are detectable. The end result is that each S1 (S2) or S1' (S2') acts as an S (S') strand, and the sum s1+s2 is measured.

Similarly, the allele difference is determined using single-stranded loops from both the upper and lower strands. This again has the advantage of signal amplification. Here, unlabelled TT' is used only as a SS-DNA protection agent, and contains no linker on its PCR primers. Instead, as with the allele difference experiment of figure 6 for determining the allele difference s2-s1, the genotyping is done by cross-hybridizing S1,S1' with S2,S2'.

Referring to figure 6, in Step 1 the STR locus and its PCR primers are chosen. In Step 2, the two complementary strands are constructed in a single PCR amplification of sample genomic DNA. In one embodiment, there are two labels on the upper strand: the first loop quantitation label (*) is

present on nucleotides (in equal proportions) in both S and S'. The second label (#) for strand quantitation is attached to the left primer. On the complementary lower strand S', there is one linker such as biotin, which is attached to the 5' end of the right primer.

The hybridization is performed in Step 3 of figure 6. Referring to figure 11, with s2>s1, the hybridization product 1102 of the denaturation and reannealing is shown in subfigure 11A. The various label and linker combinations are shown in the hybridization product table of subfigure 11B. Adding up the signals from the first label (*) 1104,

$$2n\ (S2,S1') + 2n\ (S1,S2') = 4n,$$

and adding up the signals from the second label (#) 1106,

$$1\ (S1,S1') + 1\ (S2,S2') + 1\ (S2,S1') + 1\ (S1,S2') = 4.$$

Referring to figure 6, in the detection Step 4, relative to the single-stranded detection case, there is greater signal strength from the loops and strands. The 4n loop size signal from the first label (*) represents a four-fold improvement over the single loop detection method originally described above. In Step 5, the allele difference s2-s1 is computed as n, i.e., the normalized (and calibrated) ratio of loop size signal from the first label (*) to strand number signal from the second label (#).

As in the Steps of figure 9, with appropriate linker separation and detection, the two separate sum and difference experiments can be combined into a single experiment.

D.2. Method for Genotyping STRs using Nucleic Acid Synthesis

Referring to figure 13, a method is described for determining a sum (or average) of STR alleles by nucleic acid synthesis that is comprised of the steps:

(1) Identifying an STR, and synthesizing suitable PCR reagents;

(2) PCR amplification of template DNA using the PCR reagents;

(3) Purification of amplified complementary lower DNA strand;

(4) Nucleic acid synthesis of the upper strand;

(5) Detecting signals from the synthesized nucleic acids;

(6) Analyzing the detected signals to determine the genotype sum (or average).

Referring to figure 13, step 1 is for identifying an STR, and synthesizing suitable PCR reagents.

The STR locus is identified by conventional techniques (Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning, second edition*. Plainview, NY: Cold Spring Harbor Press; N. J. Dracopoli, J. L. Haines, B. R. Korf, C. C. Morton, C. E. Seidman, J. G. Seidman, D. T. Moir, and D. Smith, ed., *Current Protocols in Human Genetics*. New York: John Wiley and Sons, 1994), incorporated by reference. Alternatively, preexisting STR loci for the genome of interest can be obtained from available databases (Genbank, GDB, EMBL; Hilliard, Davison, Doolittle, and Roderick, Jackson laboratory mouse genome database, Bar Harbor, ME; SSLP genetic map of the mouse, Map Pairs, Research Genetics,

Huntsville, AL), incorporated by reference. The STR's repeat unit includes no more than three distinct nucleotides; for clarity in exposition, the following specification of the preferred embodiment assumes that the STR is a CA-repeat marker.

The nucleic acid sequences flanking the CA-repeat region are determined by DNA sequencing methods (Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning, second edition*. Plainview, NY: Cold Spring Harbor Press; United States Biochemical 1994. USB Sequenase version 2.0 DNA sequencing kit, sequencing protocols, 9th edition, product number 70770, Amersham Life Science, Arlington Heights, IL), incorporated by reference. Alternatively, the sequence of all or part of the STR locus may reside in a preexisting available database, or in the original articles describing the locus.

Three oligonucleotide primers are designed for use with the DNA sequence using computer programs that facilitate PCR primer or DNA synthesis oligonucleotide design, such as MacVector 4.1 (Eastman Chemical Co., New Haven, CT) or Oligo 4.0 (National Biosciences, Inc., Plymouth, MN), incorporated by reference. These programs facilitate selecting lengths and positionings of oligonucleotides that are operative for enzymatic reactions. The two PCR primers and the reaction conditions are designed to permit amplification of the DNA sequence, and include:

(L) a left PCR primer for the upper strand, and

(R') a right PCR primer for the complementary lower strand. In the preferred embodiment, the 5' end of primer R' is biotinylated.

A third oligonucleotide for DNA sequencing primer and its reaction conditions are designed to permit sequencing of the DNA sequence:

(Q) a left (upstream) DNA sequencing primer that is directly adjacent to the CA-repeat region of the upper strand; this sequencing primer is designed to allow extension across the entire tandem repeat sequence using nucleotides that are specifically limited to the repeat unit base composition.

The oligonucleotide primers for the CA-repeat genetic marker are synthesized (Haralambidis, J., Duncan, L., Angus, K., and Tregear, G.W. 1990. The synthesis of polyamide-oligonucleotide conjugate molecules. *Nucleic Acids Research*, 18(3): 493-9. Nelson, P.S., Kent, M., and Muthini, S. 1992. Oligonucleotide labeling methods. 3. Direct labeling of oligonucleotides employing a novel, non-nucleosidic, 2-aminobutyl-1,3-propanediol backbone. *Nucleic Acids Research*, 20(23): 6253-9. Roget, A., Bazin, H., and Teoule, R. 1989. Synthesis and use of labelled nucleoside phosphoramidite building blocks bearing a reporter group: biotinyl, dinitrophenyl, pyrenyl and dansyl. *Nucleic Acids Research*, 17(19): 7643-51. Schubert, F., Cech, D., Reinhardt, R., and Wiesner, P. 1992. Fluorescent labelling of sequencing primers for automated oligonucleotide synthesis. *Dna Sequence*, 2(5): 273-9. Theisen, P., McCollum, C., and Andrus, A. 1992. Fluorescent dye phosphoramidite labelling of oligonucleotides. *Nucleic Acids Symposium Series*, 1992(27):

99-100.), incorporated by reference. These primers may be derivatized with a fluorescent detection molecule or a ligand for immunochemical detection such as digoxigenin. Alternatively, these oligonucleotides and their derivatives can be ordered from a commercial vendor (Research Genetics, Huntsville, AL).

Referring to figure 13, step 2 is for PCR amplification of template DNA using the PCR reagents.

A genetic material whose genotype is to be determined is selected for study. This genetic material is then placed in contact with the PCR primers 'L and 'R', and PCR amplification is performed. The methods for this PCR amplification given here are standard, and can be readily applied to every CA-repeat or microsatellite marker that corresponds to a (relatively unique) location on a genome.

In the preferred embodiment, the genomic DNA is mixed with the other components of the PCR reaction at 4°C. These other components include, but are not limited to, the standard PCR buffer (containing Tris pH8.0, 50 mM KCl, 2.5 mM magnesium chloride, albumin), triphosphate deoxynucleotides (dTTP, dCTP, dATP, dGTP), the thermostable polymerase (e.g., Taq polymerase). The total amount of this mixture is determined by the final volume of each PCR reaction (preferably 10ul to 100ul), and the number of reactions.

The PCR reactions are performed on all of the reactions by heating and cooling to specific locus-dependent temperatures that are given by the known PCR conditions. The entir cycle of annealing, extension, and denaturation is repeated multiple times (ranging from 20-40 cycles depending

on the efficiencies of the reactions and sensitivity of the detection system) (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. 1990. *PCR Protocols: A Guide to Methods and Applications*. San Diego, CA: Academic Press.), incorporated by reference. In the preferred embodiment, for STR CA-repeat loci, the thermocycling protocol on the Perkin-Elmer PCR System 9600 machine is:

a) Heat to 94°C for 3'
b) Repeat 30x:
    94°C for 1/2' (denature)
    53°C for 1/2' (anneal)
    65°C for 4' (extend)
c) 65°C for 7' (extend)
d) 4°C soak ad librum

The PCR cycles are completed, with each reaction tube containing the amplified DNA from a specific location of the genome. Each mixture includes the DNA that was synthesized from the two alleles of the diploid genome (a single allele from haploid chromosomes as is the case with the sex chromosomes in males or in instances of cells in which a portion of the chromosome has been lost such as occurs in tumors, or no alleles when both are lost). If desired, the free deoxynucleotides and primers may be separated from the PCR products by filtration using commercially available filters (Amicon, "Purification of PCR Products in Microcon Microconcentrators," Amicon, Beverly, MA, Protocol Publication 305; A. M. Krowczynska and M. B. Henderson, "Efficient Purification of PCR Products Using Ultrafiltration," *BioTechniques*, vol. 13, no. 2, pp. 286-289, 1992), incorporated by reference.

Referring to figure 13, step 3 is for purification of the amplified complementary lower DNA strand.

The lower biotinylated strand is purified from the upper strand by using magnetic streptavidin coated beads (Dynal International, Oslo, Norway). Specifically, the steps of Dynabead preparation, PCR product immobilization, DNA duplex melting using a 0.1M NaOH solution, and separation of the upper and lower DNA strands to purify the lower strand are done, as described (DYNAL 1993. Dynabeads biomagnetic separation system, Technical Handbook: Molecular Biology, Dynal International, Oslo, Norway), incorporated by reference. Specifically, with annotations:

(1) Prepare 100ul Dynabeads (excess)

Use 20 ul (200ug) of washed Dynabeads per PCR reaction.

(a) Pipette off supernatant while holding tube by magnet

(b) Wash beads x 2

- Resuspend beads in 100 ul 1x Dynabead buffer
- While holding tube near magnet, pipette off supernatant

(c) Resuspend in 200ul 2x Dynabead buffer

(Dynabeads concentration now 5 ug/ul

(2) Immobilize PCR product

Use 0.5 ug genomic DNA and 5-10 pmole of each PCR primer.

(a) add PCR product

Remove 40 ul PCR material from PCR tube under oil with pipette.

Add 40 ul Dynabead to 40ul PCR product

(b) incubate at room temp for 30 minutes
Gently rotate tube to keep Dynabeads suspended.

(3) Melting the DNA duplex
(a) pipette off supernatant while holding tube near magnet
(b) add 40ul 1x Dynabead buffer
IMMOBILIZED PRODUCT IS NOW STABLE (4oC x several weeks)
(c) pipette off supernatant while holding tube near magnet
(d) add 8 ul 0.1M NaOH solution (freshly prepared)
(e) incubate at room temp for 10 minutes

(4) Separating the DNA strands
(a) pipette off supernatant while holding tube by magnet,
and store supernatant in another tube
supernatant = nonbiotinylated strand
(b) wash x3:
- add 50 ul 0.1M NaOH solution (freshly prepared);
pipette off supernatant while holding tube near magnet
- add 40 ul 1x Dynabead buffer;
pipette off supernatant while holding tube near magnet
- add 50 ul 1x TE buffer;
pipette off supernatant while holding tube near magnet
(c) adjust volume with water for sequencing reaction
- add 7 ul sterile water

For 20ml 2x Dynabead binding and washing buffer:
16 ml 2.5 M NaCl
200 ul 1 M TRIS, pH 7.6
40 ul 0.5M EDTA
3.76 ml sterile water

Referring to figure 13, step 4 is for nucleic acid synthesis of the upper strand.

The purified amplified lower DNA strand serves as a template for a sequencing reaction. Starting from the left flanking primer Q, the sequencing reaction provides a template-directed synthesis that extends the upper strand across the CA-repeat region. The nucleotides used are:

(Q) The DNA sequencing primer that flanks the CA-repeat region, and initiates the sequencing reaction.

(dNTPs) Extension is largely restricted to the repetitive sequence by including only dNTPs that appear in the repeat unit. For a CA-repeat, only dATP and dCTP are used. One or both of these dNTPs are labeled with a detectable label *, preferably a radioisotope such as ^{35}S or ^{32}P (DuPont NEN Research Products, Boston, MA), or a fluorescent probe (Biological Detection Systems, Pittsburgh, PA). When using fluorescein-labeled dUTP (DuPont NEN Research Products, Boston, MA), the roles of the "upper" and "lower" strands are exchanged, so that the template (rather than the synthesized product) contains the CA-repeat.

(ddNTP) Termination is restricted to nucleotides not contained in the repetitive sequence. For a CA-repeat marker, ddGTP or ddTTP (ddUTP) are used, depending on the

sequence of the marker. The termination molecule is labelled with a second label **, that is distinct from the first label *, and can be independently detected. When a radioisotope is used for the first label *, fluorescein-labeled ddNTP (DuPont NEN Research Products, Boston, MA) is a convenient second label **.

Sequencing is done using standard DNA sequencing protocols (Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning, second edition*. Plainview, NY: Cold Spring Harbor Press; Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., ed. 1993. *Current Protocols in Molecular Biology*. New York, NY: John Wiley and Sons; N. J. Dracopoli, J. L. Haines, B. R. Korf, C. C. Morton, C. E. Seidman, J. G. Seidman, D. T. Moir, and D. Smith, ed., *Current Protocols in Human Genetics*. New York: John Wiley and Sons, 1994), incorporated by reference. A highly processive polymerase enzyme having little or no exonuclease activity is preferably used, such as Sequenase 2 (U.S. Biochemical, Cleveland, OH). Protocols optimized for the selected enzyme (United States Biochemical 1994. USB Sequenase version 2.0 DNA sequencing kit, sequencing protocols, 9th edition, product number 70770, Amersham Life Science, Arlington Heights, IL), incorporated by reference, are applied, with the (labeled and unlabeled) dNTPs and ddNTPs described above substituted for the dNTPs and ddNTPs contained in the conventional sequencing protocol. The use of Mn buffer can be helpful when synthesizing short sequences.

Referring to figure 13, step 5 is for detecting signals from the synthesized nucleic acids.

The newly synthesized upper DNA sequence formed by means of the DNA sequencing reaction remains hybridized to the biotinylated lower strand, which in turn is tightly bound to the streptavidin beads. The DNA sequencing primers, nucleotides, and other reagents are removed by repeated gentle washing with a buffer that promotes double stranded DNA, such as the Dynabead binding and washing buffer (DYNAL 1993. Dynabeads biomagnetic separation system, Technical Handbook: Molecular Biology, Dynal International, Oslo, Norway), leaving only the bound duplex DNA containing the desired purified product. Since the only labels present in the duplex reside on the newly synthesized upper DNA sequence (with no label * or ** present on the lower template DNA), the strands need not be separated. Fluorescence signals are detected and quantitated, preferably by means of a fluorimeter. Radioactive signals are detected and counted, preferably by means of a scintillation counter.

For quality assurance or development work, standard sequencing gels can be used for detecting signals from the synthesized nucleic acids (Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning, Second Edition*. Plainview, NY: Cold Spring Harbor Press), incorporated by reference. These protocols include a DNA denaturation step.

Referring to figure 13, step 6 is for analyzing the detected signals to determine the genotype sum (or average).

The ratio of the repeat unit label * to the end label ** varies in direct proportion to the number of tandem repeats. Precalibration with a set of predetermined reference alleles can establish the scal factor, and any deviations from linearity. PCR stutter artifact is accounted for by

deconvolution with the known stutter distribution (Perlin, M.W., Burks, M.B., Hoop, R.C., and Hoffman, E.P. 1994. Toward fully automated genotyping: allele assignment, pedigree construction, phase determination, and recombination detection in Duchenne muscular dystrophy. *Am. J. Hum. Genet.*, 55(4): 777-787), incorporated by reference.

For a single allele (e.g., hemizygote or homozygote), this analysis procedure computes the genotype. For more than one allele (e.g., heterozygote), this procedure computes the average (or, equivalently, the sum) of the alleles.

Referring to figure 14, a method is described for determining a difference of STR alleles by nucleic acid synthesis that is comprised of the steps:

(1) Identifying an STR, and synthesizing suitable PCR reagents;
(2) PCR amplification of template DNA using the PCR reagents;
(3) Purification of amplified complementary lower DNA strand;
(4') Nucleic acid synthesis of the upper strand;
(5) Detecting signals from the synthesized nucleic acids;
(6') Analyzing the detected signals to determine the genotype difference.

Steps 1, 2, 3, and 5 have been described in figure 13.

Referring to figure 14, step 4' is for nucleic acid synthesis of the upper strand, and is comprised of the steps:

(4' a) Unlabeled restricted synthesis.
(4' b) Heteroduplex formation.

(4' c) Labeled restricted synthesis.

Referring to figure 14, step 4'a is for unlabeled restricted synthesis of the upper strand.

The purified amplified lower DNA strand serves as a template for a sequencing reaction. Starting from the left flanking primer Q, the sequencing reaction provides a template-directed synthesis that extends the upper strand across the CA-repeat region. The nucleotides used are:

(Q) The DNA sequencing primer that flanks the CA-repeat region, and initiates the sequencing reaction.

(dNTPs) Extension is largely restricted to the repetitive sequence by including only dNTPs that appear in the repeat unit. For a CA-repeat, only dATP and dCTP are used. These are both unlabeled.

(ddNTP) These are specifically excluded from the reaction mixture.

Sequencing is done using standard DNA sequencing protocols (Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning, second edition*. Plainview, NY: Cold Spring Harbor Press; Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., ed. 1993. *Current Protocols in Molecular Biology*. New York, NY: John Wiley and Sons; N. J. Dracopoli, J. L. Haines, B. R. Korf, C. C. Morton, C. E. Seidman, J. G. Seidman, D. T. Moir, and D. Smith, ed., *Current Protocols in Human Genetics*. New York: John Wiley and Sons, 1994), with an excess of dNTPs relative to primer and template. A highly

processive polymerase enzyme having little or no exonuclease activity is preferably used, such as Sequenase 2 (U.S. Biochemical, Cleveland, OH). Protocols optimized for the selected enzyme (United States Biochemical 1994. USB Sequenase version 2.0 DNA sequencing kit, sequencing protocols, 9th edition, product number 70770, Amersham Life Science, Arlington Heights, IL) are applied, and the unlabeled dNTPs described above are substituted for the dNTPs and ddNTPs contained in the standard sequencing protocol. Washing with the stabilizing Dynabead binding and washing buffer is then done 2-4 times (DYNAL 1993. Dynabeads biomagnetic separation system, Technical Handbook: Molecular Biology, Dynal International, Oslo, Norway) to remove the unincorporated primers and dNTPs, and thereby purify the duplex DNA comprised of lower strand template and partially synthesized unlabeled upper strand DNA.

Referring to figure 14, step 4'b is for heteroduplex formation between different alleles of the upper and lower strands.

In the preferred embodiment, sodium hydroxide is used to melt the duplex, and an equimolar amount of hydrochloric acid is then subsequently used to reanneal (DYNAL 1993. Dynabeads biomagnetic separation system, Technical Handbook: Molecular Biology, Dynal International, Oslo, Norway). Specifically (p. 23), using the bead-immobilized double stranded product,

(3) Melting the DNA duplex

(c) pipette off supernatant while holding tube near magnet

(d) add 8 ul 0.1M NaOH solution (freshly prepared)

(e) incubate at room temp for 10 minutes

(4') Reannealing the DNA duplex

(a) neutralize with 4ul 0.2M HCl and 1ul 1M Tris-HCl (pH adjusted to optimum of sequencing enzyme).

(b) mix immediately with a pipette and adjust the volume with water according to the sequencing protocol.

(c) the same pipette is always used for both NaOH and HCl to avoid small differences in calibration that can cause neutralization problems.

In an alternative embodiment, the denaturing and renaturing is done by heating the duplex DNA solution to a temperature of 65°C to 95°C for a period of 2 to 30 minutes, and then gradually cooling the solution over a period of 15 to 90 minutes to a temperature between 25°C and 40°C.

Referring to figure 14, step 4'c is for labeled restricted synthesis of the upper strand.

The purified amplified lower DNA strand serves as a template for continuing the sequencing reaction. Starting from the left flanking primer Q that has been partially extended across the CA-repeat region, the template-directed synthesis continues the upper strand sequencing across the CA-repeat region. The nucleotides used are:

(Q) No additional DNA sequencing primer is used.

(dNTPs) Extension is largely restricted to the repetitive sequence by including only dNTPs that appear in the repeat unit. For a CA-repeat, only dATP and dCTP are used. One or both of these dNTPs are labeled with a detectable label *, preferably a radioisotope such as ^{35}S or

^{32}P (DuPont NEN Research Products, Boston, MA), or a fluorescent probe (Biological Detection Systems, Pittsburgh, PA). When using fluorescein-labeled dUTP (DuPont NEN Research Products, Boston, MA), the roles of the "upper" and "lower" strands are exchanged, so that the template (rather than the synthesized product) contains the CA-repeat.

(ddNTP) Termination is restricted to nucleotides not contained in the repetitive sequence. For a CA-repeat marker, ddGTP or ddTTP (ddUTP) are used, depending on the sequence of the marker. The termination molecule is labelled with a second label **, that is distinct from the first label *, and can be independently detected. When a radioisotope is used for the first label *, fluorescein-labeled ddNTP (DuPont NEN Research Products, Boston, MA) is a convenient second label **.

Sequencing is done using standard DNA sequencing protocols (Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning, second edition*. Plainview, NY: Cold Spring Harbor Press; Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., ed. 1993. *Current Protocols in Molecular Biology*. New York, NY: John Wiley and Sons; N. J. Dracopoli, J. L. Haines, B. R. Korf, C. C. Morton, C. E. Seidman, J. G. Seidman, D. T. Moir, and D. Smith, ed., *Current Protocols in Human Genetics*. New York: John Wiley and Sons, 1994), incorporated by reference. A highly processive polymerase enzyme having little or no exonuclease activity is preferably used, such as Sequenase 2 (U.S. Biochemical, Cleveland, OH). Protocols optimized for the selected enzyme (United States Biochemical 1994. USB Sequenase version 2.0 DNA sequencing kit, sequencing protocols, 9th edition, product number 70770,

Amersham Life Science, Arlington Heights, IL) are applied, and the (labeled and unlabeled) dNTPs and ddNTPs described above are substituted for the dNTPs and ddNTPs contained in the standard sequencing protocol.

The result of this unlabeled/heteroduplex/labeled restricted sequencing reaction is a set of four possible newly synthesized upper strands, corresponding to the two alleles s and t, where the length of allele s is less than or equal to the length of allele t:

(s,s') This homoduplex product is unlabeled with *, and may have a **-labeled terminator dye.

(t,t') This homoduplex product is unlabeled, and may have a **-labeled terminator dye.

(t,s') This heteroduplex product is unlabeled, and may have a **-labeled terminator dye.

(s,t') From the 5' end, this heteroduplex product is comprised of unlabeled primer, an unlabeled repetitive sequence with about s repeated CA units, a *-labeled repetitive sequence with about (t-s) repeated CA units, and has a **-labeled terminator dye.

Referring to figure 14, step 6' is for analyzing the detected signals to determine the genotype difference.

The ratio of the repeat unit label * to the end label ** varies in direct proportion to the number of tandem repeats. Since only one quarter of the reannealed duplexes contain **, the constant applied to the ratio of label * to label ** is greater than that of step 6 of figure 13. Precalibration with a set of predetermined reference alleles can establish this scale factor, and any deviations from linearity. PCR stutter artifact is accounted for by deconvolution with the

known stutter distribution (Perlin, M.W., Burks, M.B., Hoop, R.C., and Hoffman, E.P. 1994. Toward fully automated genotyping: allele assignment, pedigree construction, phase determination, and recombination detection in Duchenne muscular dystrophy. *Am. J. Hum. Genet.*, 55(4): 777-787), incorporated by reference.

For a pair of alleles (e.g., heterozygote), this analysis procedure computes the difference between the two alleles of the genotype.

Referring to figure 15, a method is described for determining STR alleles by nucleic acid synthesis that is comprised of the steps:

Perform the steps of figure 13.

Perform the steps of figure 14.

Combine the recalibrated ratio of label * to label ** from step 6 of figure 13, together with the recalibrated ratio of label * to label ** from step 6' of figure 14. This combination is preferably done by precalibration with a set of predetermined reference alleles that establish the mapping from the pair of measured ratios to the actual allele pairs. Alternatively, the alleles s and t are computed directly from the sum (or average) s+t and difference s-t. PCR stutter artifact is accounted for by deconvolution with the known stutter distribution (Perlin, M.W., Burks, M.B., Hoop, R.C., and Hoffman, E.P. 1994. Toward fully automated genotyping: allele assignment, pedigree construction, phase determination, and

recombination detection in Duchenne muscular dystrophy. *Am. J. Hum. Genet.*, 55(4): 777-787).

Referring to figures 13, 14, and 15, alternative embodiments for determining STR alleles by nucleic acid synthesis are given:

(a) Ligation with a reporter sequence R that flanks the CA-repeat region immediately the right (downstream) can be used, instead of a ddNTP dye terminator.

(b) Other label molecules (such as biotin) can be used on the newly synthesized upper strand. In one embodiment, the lower PCR amplified strand is constructed with a cleavable biotinylated primer (Pierce, Rockford, IL), such as a disulfide link that can be subsequently cleaved with a reducing agent (e.g., DTT). The upper strand is then synthesized from the three (5' to 3') consecutive units:

(Q) A primer that is end-labeled with the strand counter second label **.

(dNTPs) Nucleotides that are restricted to the composition of the repetitive unit, at least one of which is labeled with the repeat counter first label *. For a CA-repeat, this could be *-dATP and dCTP.

(R) A biotinylated reporter R that is added after the reducing agent has cleaved the biotinylated PCR primer from the streptavidin beads. In one embodiment, the reporter R is a biotinylated terminating ddNPT that is added by means of a sequencing enzyme. In another embodiment, reporter R is a biotinylated oligonucleotide that is added as the

right flanking sequence of the repetitive sequence by means of a ligation enzyme.

(c) The detection reagents used for the required labeling may include (but are not limited to) radioactivity, fluorescence, phosphorescence, chemiluminescence, electrical resistivity, pH, and ionic concentration.

(d) The lower strand can be sequenced, instead of the upper strand.

(e) A repetitive unit other than CA, but containing no more than three distinct nucleotides, can be used. In this case, dNTPs are used for every nucleotide in the repetitive unit, with at least one of the repetitive unit nucleotides labeled with the first label *, and ddNTP(s) are used for every nucleotide not in the repetitive unit, with the appropriate terminating nucleotide immediately following the repetitive sequence labeled with the second label **.

E. Method for Genotyping STRs using a Hybridization Panel

The hybridization panel method for genotyping STRs is distinguished from the loop mismatch method described previously in that the determination of an STR's alleles is accomplished with an entire panel of hybridization probes, rather than determining the alleles with only two loop mismatch hybridization experiments. This hybridization panel method generally entails more hybridization experiments per STR than the loop mismatch method. However, this approach is applicable to the determination of specific nucleotide

sequences realted to genomic DNA, specific genes, and known mutations.

The central idea of the hybridization panel method for genotyping STR alleles is to have a detection panel of DNA probes. When an apparatus for genotyping multiple STRs is used, each spatial location of said apparatus corresponds to one genetic STR locus and contains a separate detection panel. This panel measures the extent of specific DNA binding of the patient's DNA against a set of probes. A second coordinate of information can optionally be obtained by performing the reactions over a range of reaction stringencies (e.g., using temperature, ion concentration, or DNA denaturants). The result is a mapping from one or two coordinates (probe and stringency) into the reaction energetics (binding affinity). Different alleles produce different energy surfaces. Hence, unique pairwise combinations of alleles will produce unique signature patterns. By performing the experiment described herein, the signature can be observed, hence the zero, one, or two alleles at a sample point uniquely determined.

E.1. Method for Genotyping STRs using a Direct Hybridization Panel

To fix ideas, let $L(CA)_nR$ be one allele in the patient's PCR product for a given STR reaction chamber in the two dimensional array. Here, L is the left flanking region DNA subsequence, R is the right flanking region DNA subsequence, and n is the number of allelically varying CA repeats, so that $(CA)_n$ is the middle DNA subsequence of length 2n. The left PCR primer (denoted by P) is a prefix subsequence of the

left flanking region L, and the right PCR primer (denoted by S) is a suffix subsequence of the right flanking region R. For constructing probes to such PCR products, note that a GT polymer binds complementarily to a CA polymer.

In a preferred embodiment for constructing said detection panel, each detection panel is customized to the PCR product of its STR allele. This is done by providing a panel of allele specific oligonucleotides (ASOs) (Lemna, W.K., Feldman, G.L., Kerem, B.-S., Fernbach, S.D., Zevkovich, E.P., O'Brien, W.E., Riordan, J.R., Collins, F.S., Tsui, L.-C., and Beaudet, A.L. 1990. Mutation analysis for heterozygote detection and the prenatal diagnosis of cystic fibrosis. *N. E. J. Med.*, 322: 291-296), incorporated by reference, where each ASO contains an allele-specific left flanking region, concatentated with a number n of repeat unit nucleotides, concatentated with an allele-specific right flanking region. The lengths of the left and right regions flanking the varying size repeat polymer are individually adjusted to ensure that the left and right oligomers have roughly the same DNA binding energies when hybridizing to their respective complementary DNA strands.

The thermodynamic basis for this (and alternative) approaches is that while perfect DNA duplex matches will have minimum energy, mismatches will induce bulges or loops in the DNA duplex molecule that increase the free energy. A two base-pair bulge will have sufficiently increased free energy (Ninio, J. 1979. *Biochimie*, 61: 1133. Salser 1977. *Cold Spring Harbor Symp. Quant. Biol.*, 42: 985.), incorporated by reference, to reduce binding affinity by several kcal/mole relative to a perfect match; the larger the bulge, the more unfavorable the binding. Therefore, given a STR target with

n repeating units in the middle (anchored by left and right flanking sequences), and a STR source PCR product with m of complementary repetitive units (anchored by the complementary left and right flanking sequences), high stringency DNA hybridization is a sensitive measure of whether or not m=n. In this way, a panel of ASOs that provide for all values of n is used to determine the m values expressed from the PCR product.

With CA-repeats as STRs, each DNA target probe in the panel has the form $L_0(CA)_nR_0$, or the complementary form $L_0'(CA)_n'R_0'$, where n varies across the polymorphic $(CA)_n$ alleles of the genetic locus (say, n = 15, 16, ..., 30), L_0 is a suffix of the DNA flanking sequence L, R_0 is a prefix of the DNA flanking sequence R, and U' is the complementary strand of DNA sequence U.

Consider the example panel of target probes for the STR-45 locus residing in an intron of the dystrophin gene (Clemens, P., Fenwick, R., Chamberlain, J., Gibbs, R., de Andrade, M., Chakraborty, R., and Caskey, C. 1991. Linkage analysis for Duchenne and Becker muscular dystrophies using dinucleotide repeat polymorphisms. *Am J Hum Genet*, 49: 951-960.), incorporated by reference. 15 bases are taken from the left flanking AT-rich region, and 10 bases from the right flanking GC-rich region in order to equalize the DNA hybridization energies, as

L_0 = ATTAGTTGACCTAAA

R_0 = CCCCTTGCCA

Target probes are then constructed by inserting $(CA)_n$ units, e.g.,

$(CA)_{10}$= CACACACACACACACACACA.

Then, the panel of target probes is constructed as the set of DNA sequences formed by concatenating L_0, $(CA)_n$, and R_0, as

{ L_0 $(CA)_n$ R_0 | n varies from 10 to 40 by 1 }.

The complementary PCR source products have the form

{ L_0' $(GT)_m'$ R_0' | m varies from 10 to 40 by 1 }.

When an exact match occurs between allele source and probe target, i.e., the GT-repeat polymer length m exactly equals the CA-repeat polymer n, the binding is energetically most favorable (i.e., stable). Thus, under appropriate hybridization binding conditions (Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning, second edition*. Plainview, NY: Cold Spring Harbor Press.), incorporated by reference, the two alleles m1 and m2 will bind most avidly to the two probes in the target panel having the corresponding n1=m1 and n2=m2. The detection of the two specific targets n1 and n2 out of the entire target panel can be effected by a variety of methods, as described next.

Confirmation of the energetics for the STR-45 locus target panel can be seen in the following data generated by running Zuker's RNA folding program (Zuker, M., and Stiegler, P. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids*

Research, 9: 133-148.), incorporated by reference. The left and right flanking sequences each contain ten bases. The temperature here is set to 70°C, a source is used with m = 21, and a panel of targets with n = 18, 19, ..., 24. As shown, the energetic difference between target 21 and its nearest neighbors exceeds 2 kcal/mole, and is thus unamibiguously detectable.

Target	18	19	20	21	22	23	24
kcal/mole	-45.4	-48.4	-51.7	-57.5	-53.8	-52.6	-51.7

To implement this differential detection, one detection panel is provided for the PCR products of each genetic marker. Each detection panel corresponds to one marker locus, and is embedded at that locus' coordinate in the spatially localized PCR marker grid. The two surfaces (PCR and detection) may be separate or composite. In this detection panel scheme, the oligomers flanking the STR region are (in general) different for every genetic marker. That is, the target probe panel sequences are customized to each genetic marker.

In another preferred embodiment, a second coordinate of hybridization stringency would be added. This stringency variation can be implemented by varying any of several factors in the hybridization, including temperature, ion concentration, formamide concentration, and nucleotide composition (Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning, second edition*. Plainview, NY: Cold Spring Harbor Press.), incorporated by reference. The two coordinates of differential targets and differential stringency give an even clearer signature for STR alleles.

The signature of two alleles is formed by superimposing those of single alleles. By predetermining all possible single allele and paired allele patterns, unique signatures (in one or two coordinates) can be generated, and then later retrieved to effect genotyping into the component alleles. This is done by comparing the measured genotype signature at a genetic locus with the retrieved signatures, and determining a best match. Alternatively, the separation of the superimposed patterns to effect genotyping can be done without recourse to such a library of signatures by curve fitting or deconvolution processing.

Such signatures are seen in simulations with Zuker's program using CA-repeats as STRs, where the parameters are as before, but, additionally, the temperature assumes the multiple values 60°C, 70°C, and 80°C. With the flanking markers, this serves to reinforce the pattern of best match when m=n.

Target	18	19	20	21	22	23	24
60°C	-61.4	-65.0	-68.9	-75.4	-71.6	-70.4	-69.5
70°C	-45.4	-48.4	-51.7	-57.5	-53.8	-52.6	-51.7
80°C	-32.0	-33.6	-36.5	-41.7	-38.1	-36.8	-36.8

In this second differential detection approach, again one unique detection panel is provided for the PCR products of each genetic marker. However, for each STR locus, the target panel is replicated for every measured stringency.

This replication can be accomplished by providing for the PCR products of the STR:

(1) A single panel, reused at different times with varying stringencies. The stringency variation can be effected by temperature ramp, or by changing the chemical environment of the hybridization over time.

(2) Multiple panels, usable at the same or different times, with varying stringencies. Here, the genetic locus grid and its PCR amplification is replicated across each of the multiple target panels.

(3) Multiple panels on the same surface. This is done by placing multiple target panels, each with a different stringency, on the same surface. These are all be located in the same region of one genetic locus and its PCR amplification. Alternatively, one genetic locus may be replicated multiple times on the same surface, at each position having a target panel of identical composition, but different stringency.

(4) Any combination of the above.

An alternative embodiment uses an identical detection panel of target oligonucleotides for every genetic locus. This has the utility of reducing manufacturing costs, since no STR locus customization is required, and the same detection panel design and manufacture is reusable for every genetic locus. With CA-repeats as STRs, each grid is comprised of the target panel

{ $(CA)_n$ | n varies across all interesting polymorphisms }.

For example, n could range from 10 to 40.

In another embodiment, intentional DNA pairing mismatch is introduced to bias the hybridization against further STRs. This can be done by a three-fold expansion of these probes by adding a mismatching base pair at one end. For example, with CA-repeats as the STR, these four probe families are possible for every n:

$\{A,C,T\}(GT)_n$, $\{C,G,T\}(CA)_n$, $(GT)_n\{A,C,T\}$, or $(CA)_n\{A,G,T\}$.

Within each family, say $\{A,G,T\}(CA)_n$, the three probes

$C(CA)_n$, $G(CA)_n$, $T(CA)_n$, but not $A(CA)_n$,

are provided. The idea is that an intentional mismatch is introduced to avoid the close energetics produced from DNA slippage during hybridization.

Extending this, there is a nine-fold expansion of the STR by introducing intentional mismatch on both sides of the repeat region. For example, with CA-repeats as the STR, these nine probes are generated for every n:

$\{C,G,T\}(CA)_n\{A,G,T\}$.

This provides a better balanced mismatch. The strategy has been used in PCR primer design for developing

microsatellite markers. This approach can be further extended to introduce as much bias against STR extension as desired, by building targets for every n that have some number of STR blockers on the left, and some number of STR blockers on the right. The main advantage of this intentional mismatch approach is improved STR specificity for a fixed length. The main disadvantage is the increased number of target DNA probes required in the detection panel.

In another embodiment, the same detection panel is used for every genetic locus, but intentional mismatch is introduced by changing the target DNA composition. With CA-repeats as STRs, a family of $(CA)_n$ or $(GT)_n$ probes are used, but changes are introduced in specific bases. For example, some G's are changed to C's, or to the energetically similar base inosine. One (of many) doping strategy is to introduce k evenly spaced doping sites, where k = 0,1,2, and so on. In general, doping the targets reduces binding affinities in a selective way.

In another embodiment, the doping is introduced in the source molecule, rather than in the targets. This has the advantage of requiring just one target DNA molecule (i.e., a very large repeated oligomer) for all the genetic loci. Thus, the manufacturing costs are greatly reduced, since replicated complex panels for each locus are not needed. The extent of doping is introduced (say, with inosine) as a variable into the PCR reaction itself. The doping is random across the PCR products, but has constant statistics, particularly in the repetitive unit region of the unknown STR PCR product molecule. If two coordinate signatures are

desired, hybridization stringency variation can be introduced as well.

In another embodiment, a single STR detection probe is used for all experiments. Using a single probe, say $(CA)_n$ (n large and fixed), dramatically reduces manufacturing costs. A temperature ramp experiment is then conducted in parallel for every genetic locus by varying stringency. For each PCR product with GT-repeat length, when its subpopulation of $(GT)_k$ sequences rapidly melts, there will be a sharp change in the melting profile. This will be detectable as a peak in the first derivative of the curve. The peaks provide a DNA size vs. concentration mapping that can then be used to determine the alleles.

These embodiments work with STR repeat units of any size. The newer trinucleotide repeats, tetranucleotide repeats, etc. are more favorable energetically, and provide greater allele differentiation. In instances where uniqu DNA sequences are assayed, the size of the bound detection oligonucleotide is adjusted to maximally discriminate between a perfect match and a single base pair mismatch. An alternative to detecting perfect vs. mismatched heteroduplexes is using chemical modification reagents (such as CII, CAA, OsO_4, or hydroxylamine) that can react with single nucleotide mismatches and then be detected.

In the hybridization detections, the roles of the upper strand and the lower strand may be interchanged. With CA-repeats, this would mean that the CA-strand and the GT-strand relations would be interchanged. Nested PCR (Yourno 1992. A Method for Nested PCR with Single Closed Reaction Tubes. *PCR*

Meth. Appl., 2(1): 60-65. Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. 1990. *PCR Protocols: A Guide to Methods and Applications*. San Diego, CA: Academic Press.), incorporated by reference, can be done for a purer PCR amplification to reduce noise. Two primer pairs are used: one pair for the initial amplification, and one labeled pair for the secondary amplificiation and detection. Ligase chain reaction (LCR) (Landegren, U., Kaiser, R., Sanders, J., and Hood, L. 1988. A ligase-mediated gene detection technique. *Science*, 241: 1077-1080.), incorporated by reference, can be used in place of PCR when an assay for exact match is desired, as is the case with the described panel hybridizations.

In the hybridization detection assays described, both strands must be nucleic acids. Whether these are comprised of DNA, RNA, or any other nucleic acid polymer is nonessential. The key requirement is the binding specifity of complete and partial sequence matches. Further, these nucleic acids are modified (e.g, with linker molecules, biotin, detection moieties) to perform the detection components of the method.

E.2. Method for Genotyping STRs using a Nucleic Acid Ligation

Referring to figure 16, a schematic representation is shown of an assay for determining STR alleles from a nucleic acid ligation step.

Standard oligonucleotide ligation assay (OLA) assays for the exact match of a pair of oligonucleotides X and Y against a DNA template molecule previously amplified by PCR (Landegren, U., Kaiser, R., Sanders, J., and Hood, L. 1988.

A ligase-mediated gene detection technique. *Science*, 241: 1077-1080; Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. 1990. *PCR Protocols: A Guide to Methods and Applications*. San Diego, CA: Academic Press), incorporated by reference. Following amplification with the PCR primers L and R', two ligation oligonucleotides are conventionally used:

(X) initiates the matching sequence from the 5' end, and is biotinylated;

(Y) completes the matching sequence to the 3' end, and is labeled (e.g., with radiolabel or fluorescent label). The 5' end of Y is phosphorylated to allow ligation to X.

When the sequence XY is complementary to a subsequence of the template DNA, ligation occurs and the match is detected. For CA-repeat (or any other polynucleotide repeat) marker detection, the variable length repeat precludes the described use of this assay. However, by introducing a set of third oligonucleotides {Zk}, where each Zk is a k-fold repeat of the unit Z (Z="CA" in the preferred embodiment), CA-repeat alleles can be detected. Specifically,

(Zk) bridges the gap between X and Y. The 5' end of Zk is phosphorylated to allow ligation to X. The phosphorylated Y, in turn, is ligated to Zk.

This CA-repeat detection differs from conventional ligation assays in that (a) a three-way ligation is performed, (b) a set of intermediate molecules is used, (c) these intermediate molecules are universally reusable f r

assaying more than one CA-repeat marker, and (d) a sequence of varying length can be detected.

A panel of assays is constructed, one for each intermediate sequence Zk which has k repeats of the base unit Z. The choice of k's panel corresponds to the allele distribution (hence repeat sizes) of the CA-repeat marker. When detecting two alleles, the best Zk's which have the strongest signals determine the alleles. This detection can be improved on by deconvolving the panel of signals with the known PCR stutter pattern of the alleles (Perlin, M.W., Burks, M.B., Hoop, R.C., and Hoffman, E.P. 1994: Toward fully automated genotyping: allele assignment, pedigree construction, phase determination, and recombination detection in Duchenne muscular dystrophy. *Am. J. Hum. Genet.*, 55(4): 777-787), incorporate by reference. Deconvolution methods can be similarly applied for assaying more than two alleles, as is done in population studies.

In an alternative embodiment, ligation chain reaction (LCR) is performed, rather than a PCR amplification followed by an OLA detection step. This embodiment uses the three oligonucleotides X, Y, and Z described above. Specific protocols can be found in (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., ed. 1993. *Current Protocols in Molecular Biology*. New York, NY: John Wiley and Sons; Dracopoli, N.J., Haines, J.L., Korf, B.R., Morton, C.C., Seidman, C.E., Seidman, J.G., Moir, D.T., and Smith, D., ed. 1994. *Current Protocols in Human Genetics*. New York: John Wiley and Sons; Landegren, U., Kaiser, R., Sanders, J., and Hood, L. 1988. A ligase-mediated gene detection technique. *Science*, 241: 1077-1080), incorporated by reference.

E.3. Method for Genotyping STRs using a Nucleic Acid Loop Ligation

Referring to figure 17, a schematic representation is shown of an assay for determining STR alleles from a nucleic acid loop ligation step.

Two unique primers for a specific microsatellite are constructed. The primers are selected to flank the tandem repeat but to leave at least 15 to 20 bp of internal unique sequence flanking the repeat region.

A loop oligonucleotide is constructed from the internal, unique flanking sequences within the PCR'd product. The oligonucleotide is designed to have significant base mismatching if there is "slippage" and a portion of the oligonucleotide extends into the 5' and 3' portions of the tandem repeat. The degree of extension into the repeat can be varied but is done so that the bridging oligonucleotides are smaller, preferably 15-20 nucleotides than the loop oligonucleotide. A melting temperature for the loop oligonucleotide that is about 10° higher than the largest bridging oligonucleotide is desirable. In the preferred embodiment, the loop oligonucleotide is biotinylated or covalently bound to a support matrix or surface. In another preferred embodiment described herein, the loop oligonucleotide is bound to paramagnetic beads that are covalently linked to strepavidin. The loop oligonucleotide is phosphorylated at the 5' end.

The microsatellite marker is amplified using standard PCR primers and conditions. The double-stranded DNA is denatured and annealed to the loop oligonucleotide. The

conditions of the annealing are such that the concentrations of the DNA and oligonucleotide are relatively low to discourage concatamer formation, the loop oligonucleotide should be present in excess with respect to the PCR product. The hybridization is performed at a sufficient temperature (preferably 37°C) in 0.1xSSC or a comparable buffer such that the annealed loop oligonucleotide and PCR strand are stable, but simple annealing within the tandem repeat of the two PCR DNA strands is disfavored. The annealing is performed at a low concentration in a minimum volume of 200 microliters in order to disfavor concatamer formation.

Referring to figure 17, part A, the original PCR primers do not need to be removed prior to the annealing. After the annealing is completed, the unhybridized DNA and primers are eliminated by washing with the hybridization buffer.

Referring to figure 17, part B, both specificity and sensitivity is achieved by hybridizing the PCR product with the loop oligonucleotide. After removal of the complementary PCR product DNA strand and primers, the structure is annealed (in a set of separate chambers or positions) with a set of bridging oligonucleotides that represent different multiples of the tandem repeat. The bridging oligonucleotide is complementary to the PCR'd DNA strand that is hybridized to the loop oligonucleotide. The bridging oligonucleotide is labeled with radioactivity or another detection tag such as fluorescein. The bridging oligonucleotide is phosphorylated at the 5' end.

The exonuclease reaction is carried out to digest all noncircularized, single- or double-stranded DNAs and primers. The remaining material on the support matrix represents the

undigested circularized loop oligonucleotide and bridging oligonucleotide.

Bridging oligonucleotides that are too short or too long to perfectly close the loop oligonucleotide are ligated to one end of the loop oligonucleotide but cannot allow the structure to circularize. These partially ligated products are then eliminated during the exonuclease step.

The following ligation protocol steps are essentially as in (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. 1990. *PCR Protocols: A Guide to Methods and Applications*. San Diego, CA: Academic Press), incorporated by reference.

(1) Combine:

3 μl of PCR'd sample
1 μl of sheared salmon sperm DNA at 10 μg/ml
2 μl of H_2O

(2) Denature the above DNA by heating at 95°C for 2 minutes.

Alternatively, use alkali denaturation by replacing the 2 μl of H_2O with 1 μl of 0.5 N NaOH (room temperature for 10 minutes) followed by 1 μl of 0.5 N HCl.

(3) Add:

1 μl of 140 fmol of biotinylated loop oligonucleotide (phosphorylated)
1 μl of 1.4 fmol of bridge oligonucleotide (phosphorylated with ^{32}P)

2 μl of 0.1-0.2 Weiss units of T4 DNA ligase in 5x ligase buffer (250 mM Tris-Cl (pH=7.5), 500 mM NaCl, 50 mM $MgCl_2$, 25 mM dithiothreitol, 5 mM ATP, 500 μg/μl BSA)

(4) Terminate the reaction by heating at 95°C for 2 minutes.

After the ligation, 0.1 to 0.5 units each of exonuclease VII (which digests single-stranded DNA from 5' and 3' ends) and exonuclease III (which digests double-stranded DNA, but not single-stranded DNA). Digestion proceeds at 37°C for 30 minutes.

The nondegraded products (the circularized strands) are bound to the streptavidin-paramagnetic beads in a 500 μl tube, washed three times with 200 μl of washing buffer and then counted directly or denatured off of the beads using the loading buffer/Dye for sequencing gels and run on a standard denaturing sequencing gel.

In an alternative embodiment, the annealing and ligation of the bridge and loop oligonucleotides to create a circular structure is performed as a two-stage process to discourage concatemer formation. In this protocol, only the bridge oligonucleotide is phosphorylated. The reaction is identical to that described until the end of the ligation step. At that point, the sample is denatured at 95°C for 5 minutes and 0.1 unit of T4 Polynucleotide kinase is added at 37°C for 30 minutes. This phosphorylates the 5' ends of the loop oligonucleotides. The reaction is then again heated at 95°C for 2-5 minutes and the samples are diluted 100 fold in 1x ligase buffer to promote circularization. The diluted sample

is concentrated using the streptavidin-paramagnetic beads and then treated as above with exonucleases III and VII.

F. Method for Identifying Inheritance Patterns Using Concordance Analysis

As a disease gene segregates within a pedigree, individuals inheriting the linear chromosomal segment that contains the founder's affected disease gene will carry the disease. Chromosomal regions that are closer to the disease gene will be more tightly linked, and these regions and their associated genetic markers, will have a greater tendency to be associated with the disease. Conversely, regions and markers that are further away will be less likely to have the disease association. In an X-linked disease that is fully penetrant in males, the presence of phenotypic diseas indicates inheritance of the affected disease gene, while absence (in males) indicates that the affected disease region has not been inherited. In autosomal diseases, the unaffecteds are less useful.

Inner Product Mapping (IPM) (Perlin, 1993) is a method for mapping large physical DNA probes (e.g., > 25,000bp cosmids or YACs) that uses radiation hybrids (RHs). For each RH, a dense sampling across the chromosome (or whole genome) is first obtained using sequence tagged sites or fluorescence *in situ* hybridization. This sampling maps the regions in which large chromosomal fragments have been retained, and where they have been lost, indicated by a + or -, respectively. Additionally, every physical probe has its own signature of +'s and -'s, one for each RH, which indicates whether or not the probe lies within some fragment of the RH. The probe's RH signature is compared with the RH signature of every STS. When the signatures match at some RH (i.e., ++ or

--), this indicates concordance between the two signatures, whereas when there is a mismatch (i.e., +- or -+), this indicates discordance between the signatures. For every STS sample point along the chromosome, the sum of the matches minus the sum of the mismatches is computed, which generates a profile curve across the chromosome. The peak of this profile suggests the location of the probe. A feature of IPM is its ability to map accurately using few experiments: a logarithmic number of RHs provides linear resolving power.

Recombination events in meiosis cause the founders' chromosomal regions to be retained or lost in progeny. One can consider the chromosomal segment containing the affected disease gene as a probe. The location of this probe is suggested by the concordance of chromosomal regions that affected (or carrier) individuals share with founder(s) (++), or those regions which unaffected individuals do not share with founder(s) (--). Conversely, discordance is suggested in those chromosomal regions affected (or carrier) individuals do not share with the founder(s) (+-), or those regions which unaffected individuals share with founder(s) (-+). This motivates the application of IPM to disease gene localization.

Referring to figure 12, in Step 1 phenotypic information is obtained on a set of related individuals. In Step 2, a dense genotyping across a chromosome using highly-polymorphic STSs is obtained for all informative pedigree members; in the preferred embodiment, this is done with the apparatus of figure 1. Using phase known genotypes, haplotyping is done wherever possible. The founder genotype is obtained directly from the founder (if available), or constructed indirectly as

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/01395

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68; C12P 19/34; G01N 21/00
US CL :435/6, 91.2; 422/68.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.2; 422/68.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P --- A,P	US, A, 5,348,853 (WANG ET AL.) 20 September 1994, see Figure 15.	1-7 and 11 --- 8-10 and 12-14
A,P	US, A, 5,364,759 (CASKEY ET AL.) 15 November 1994, see entire document.	1-14
A	US, A, 5,126,239 (LIVAK ET AL.) 30 June 1992, see entire document.	1-14
A	Genomics, Volume 2, issued 1988, Skolnick et al., "Simultaneous analysis of multiple polymorphic loci using amplified sequence polymorphisms (ASPs)", pages 273-279, see entire document.	1-14

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 07 APRIL 1995	Date of mailing of the international search report 21APR1995
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer KENNETH R. HORLICK Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/01395

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, f the relevant passages	Relevant to claim N .
A	Nature, Volume 359, issued 29 October 1992, Weissenbach et al., "A second-generation linkage map of the human genome", pages 794-801, see entire document.	1-14
A	Science, Volume 245, issued 29 September 1989, Olson et al., "A common language for physical mapping of the human genome", pages 1434-1435, see entire document.	1-14

INTERNATIONAL SEARCH REPORT

Int. ..national application No.
PCT/US95/01395

Box I Observations where certain claims were found unsearchable (Continuation f item 1 of first sheet)

This international report has not been established in respect f certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ Th additional search fees were accompanied by the applicant's protest.

☒ No protest accompanied the payment of additi nal search fees.

INTERNATIONAL SEARCH REPORT

Inter.national application No.
PCT/US95/01395

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, BIOSIS, MEDLINE
Search terms: polymorphic genetic markers, genome sampling, genotyping, hybridization, high throughput, STS, sequence tagged site, VNTR

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

I. Claims 1-10, drawn to an apparatus for analyzing genetic material of an organism and methods of use thereof.

II. Claim 11, drawn to a method of manufacturing an apparatus for analyzing genetic material of an organism.

III. Claims 12-14, drawn to methods of determining the size of nucleotide sequences of an STR marker contained on genetic material.

PCT Rule 13.2(ii) provides for grouping claims to a process and an apparatus or means for carrying out the said process; thus the apparatus of claims 1-7 and the methods of claims 8-10 are grouped together. PCT Rule 13.2 does not provide for the additional grouping of claims drawn to a method of manufacturing said apparatus (claim 11) nor to a distinct method (claims 12-14). Accordingly, the claims are not so linked by a special technical feature within the meaning of PCT Rule 13.2 so as to form a single inventive concept.

the union of alleles at each locus for every carrier or affected child of the founder.

Referring to figure 12, in Step 3 let v(i) be the sign of the phenotype of an individual i, where

v(i) = +1, when i is affected or a carrier, and
= -1, when i is not affected.

Let the triple <i,m,a> denote that individual i at marker m has allele a. Genotyping over a pedigree constructs a set of such triples. In Step 4 compute

w(i,m,a) = the weight accorded the triple, as follows.

In one IPM approach, assume that the alleles are sufficiently informative for an identity-by-state (IBS) analysis. Then whether or not an individual's allele is identical to a founder's allele would be known unambiguously. Therefore, in this case, define

w(i,m,a) = +1, when the founder has allele a occurring at marker m. = -1, when allele a is not shared with the founder.

In a second IPM approach, the w(i,m,a) term weights for the probability that an allele a was transmitted to individual i at marker m by the founder. That is, an accounting for identity-by-descent (IBD) is done. At each link in the inheritance graph, the probability of descent at a marker from the founder for an allele on the chromosome is computed. The product of these link probabilities over every link in the inheritance path therefore provides an estimat

of the probability of descent. Linearly rescaling this descent probability from the range [0,1] by the function

$$f(x) = 1 - 2x$$

provides a number in the range [-1,+1], which is useful for calculations.

Under either IPM approach (IBS or IBD), in Step 5 a concordance is computed for every allele of every STS marker by summing over the individuals {i} chromosomes as

$$c(m,a) = \text{SUM (over i) } [\ v(i) * w(i,m,a)\ ].$$

Each summand is a number between -1 and +1. A marker which has an allele maximizing this sum has the greatest concordance with the founder, and suggests a chromosomal region containing the gene. Taking the maximum value c(m,a) of the alleles {a} at each marker m, in Step 6 the concordance function

$$C(m) = \text{MAX (over a) } [\ c(m,a)\ ]$$

is computed. Note that this computation proceeds directly from the allele data, and requires no analysis of recombination breakpoints.

In Step 7, the genetic regions correlating with the trait are localized. With densely sampled markers {m} at previously determined map locations, the concordance function C(m) computes a profile over the chromosome. Where this profile shows a pattern on the chromosome that rises up to a peak, and then again descends from it, suggests the location of the gene (near the peak). With autosomal or nonfully

penetrant disorders, the unaffected individuals are weighted to have less influence. While two-point or multi-point likelihood analyses are alternative embodiments to the one-point IPM approach, their algorithmic complexity may preclude practical application to dense genotyping of very many individuals. Multigenic traits will produce patterns of multiple peaks; each peak corresponds to a region on the genome that influences the trait.

Dense genotypes are obtained for related sets of individuals; in the preferred embodiment, this is done with the apparatus of figure 1. In Step 8 of figure 12, the genetic patterns obtained in Step 7 are used to assess the risk of individuals for various traits and diseases. In Step 9, the localization of disease genes on a genetic map is used to initiate the cloning of the gene via positional cloning techniques (Kerem, B.-S., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M., and Tsui, L.-C. 1989. Identification of the cystic fibrosis gene: genetic analysis. *Science*, 245: 1073-1080.Riordan, J.R., Rommens, J.M., Kerem, B.-S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M.L., Iannuzzi, M.C., Collins, F.S., and Tsui, L.-C. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, 245: 1066-1073.), incorporated by reference.

G. Useful Applications of the System

Use of the apparatus described in figure 1 with th system in figure 3 is made for health risk assessment, as described above. Dense genotyping has application to prenatal gen tic screening (Schwartz, L.S., Tarleton, J., Popovich, B., Seltzer, W.K., and Hoffman, E.P. 1992. Fluorescent

Multiplex Linkage Analysis and Carrier Detection for Duchenne/Becker Muscular Dystrophy. *Am. J. Hum. Genet.*, 51: 721-729.), incorporated by reference, and in detecting chromosomal abnormalities. Such genotyping can be used for actuarial analysis of health risks in order to predict and reduce health care costs. Genotyping also finds application in transplantation (Scharf, S., Saiki, R., and Ehrlich, H. 1988. New methodology for HLA class II oligonucleotide typing using polymerase chain reaction (PCR) amplification. *Hum. Immunol.*, 23: 143.), incorporated by reference, and in the screening and evaluation of military personnel. The loop mismatch methods described can detect exon repeats that correlate with disease and prognosis, as well as exon alleles (via multiple chemical modification assays) for precise molecular diagnostics (Beggs, A., and Kunkel, L. 1990. A polymorphic CACA repeat in the 3' untranslated region of dystrophin. *Nucleic Acids Res*, 18: 1931. Beggs, A.H., Koenig, M., Boyce, F.M., and Kunkel, L.M. 1990. Detection of 98% of DMD/BMD gene deletions by polymerase chain reaction. *Hum. Genet.*, 86: 45-48), incorporated by reference. The hybridization panel methods can similarly detect exon alleles.

The apparatus and system is useful for the positional cloning of genes that cause traits and diseases. Linkage (Ott, J. 1991. *Analysis of Human Genetic Linkage, Revised Edition*. Baltimore, Maryland: The Johns Hopkins University Press.), incorporated by reference, and other analyses use dense genotypes to elicit patterns of inheritance and localized genetic regions of influence that correlate with genes. Such patterns are useful in genetic design applications, such as animal and plant husbandry, for example, for crop improvement (Bernatzky, R. (1993). Genetic

mapping and protein product diversity of the self-incompatibility locus in wild tomato (Lycopersicon peruvianum). *Biochemical Genetics*, 31(3-4): 173-84. Ho, J.Y., Weide, R., Ma, H.M., van, W.M., Lambert, K.N., Koornneef, M., Zabel, P., and Williamson, V.M. (1992). The root-knot nematode resistance gene (Mi) in tomato: construction of a molecular linkage map and identification of dominant cDNA markers in resistant genotypes. *Plant Journal*, 2(6): 971-82.), incorporated by reference, and cataloguing strains.

Dense genotyping can be used to detect the occurrence of chromosomal patterns in a population. This applies in law enforcement applications (Jeffreys, A.J., Brookfield, J.F.Y., and Semeonoff, R. 1985. Positive identification of an immigration test-case using human DNA fingerprints. *Nature*, 317: 818-819.), incorporated by reference, for genetically fingerprinting individuals, as well in paternity testing to assess parenthood.

Genotyping can monitor the changes in the chromosomal patterns of populations, including:

- Cancer testing and assessment (Zhang, Y., Coyne, M.Y., Will, S.G., Levenson, C.H., and Kawasaki, E.S. (1991). Single-base mutational analysis of cancer and genetic diseases using membrane bound modified oligonucleotides. *Nucleic Acids Research*, 19(14): 3929-33.), incorporated by reference, determining the metastatic extent of tumor, and its sensitivity to treatment.

• *In vitro* assays for toxic, mutagenic, and other pharmacological effects of chemicals (e.g., on tissue cultures).

• The relatedness of populations, and quantitating environmental impact on populations (Atlas, et al, 1992. Molecular Approaches for Enviromental Monitoring of Microorgansisms. *BioTechniques*, 12(5): 706-714. Bej, and Mahbubani 1992. PCR Meth. Appl. *Applications of the Polymerase Chain Reaction in Environmental Microbiology*, 1(3): 151-159), incorporated by reference.

• Determining geographical spread for animal migration (e.g., fisheries) and pathogen spread (e.g., epidemiology).

• In the pest control industry for determining tolerance and susceptibility, and detecting resistance to pest control agents.

• With microorganisms (including yeast and bacteria) to characterize exon DNA for pathogenicity, or to determine causative organisms for infections (Lerman, L.S., ed. 1986. *DNA Probes: Applications in Genetic and Infectious Disease and Cancer*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory), incorporated by reference.

Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

WHAT IS CLAIMED IS:

1. An apparatus for analyzing genetic material of an organism comprising:

means for amplifying the genetic material of the organism; and

means for characterizing the amplified genetic material, said characterizing means in communication with the amplifying means, said characterizing means containing all of the genetic material within a region having a radius of less than two feet, said amplifying means and characterizing means characterizing the genetic material at a rate exceeding 100 sequence-tagged sites per hour per organism.

2. An apparatus as described in Claim 1 wherein the genetic material includes nucleotide sequences and wherein the amplifying means includes a reaction plate with which the genetic material is in contact, said reaction plate having a plurality of chambers, each of which is disposed in a unique location of the plate corresponding to a location within a genome having at least one nucleotide sequence.

3. An apparatus as described in Claim 2 wherein the characterizing means includes means for detecting whether a chamber contains a nucleotide sequence of the genetic material corresponding to the chamber's unique location.

4. An apparatus as described in Claim 3 including a thermocycler in thermal communication with the plate to heat and cool the plate.

5. An apparatus as described in Claim 4 wherein the detecting means includes a detector connected to the chambers which produces a chamber signal for each chamber corresponding to genetic material in each chamber, and a processor in communication with the detector which receives the signals and identifies unique properties of the nucleotides in each chamber.

6. An apparatus as described in Claim 5 wherein the unique properties of the nucleotide of the genetic material in each chamber pertain to a number of nucleotides in any of the nucleotide sequences of the genetic material.

7. An apparatus as described in Claim 6 wherein the amplifying means includes at least one nucleotide sequence that corresponds to each chamber in contact with the chamber, each nucleotide sequence interacting with the nucleotide sequence of the genetic material of the nucleotide sequence if it is present.

8. A method for analyzing genetic material of an organism comprising the steps of:

amplifying the genetic material; and

characterizing the amplified genetic material in a region having a radius of less than two feet at a rate exceeding 100 sequence-tagged sites per hour per organism.

9. A method as described in Claim 8 wherein the genetic material includes DNA or RNA.

10. A method as described in Claim 9 including after the characterizing step, there is the step of assessing risk of illness for which there is a genetic susceptibility in the organism.

11. A method for manufacturing an apparatus for analyzing genetic material of an organism comprising the steps of:

placing corresponding sequence-tagged sites in contact with corresponding chambers of a plate;

connecting detectors to the chambers which can detect whether nucleotide sequences of the genetic material of the organism, when placed in contact with the chambers, have reacted with the corresponding sequence-tagged sites in the corresponding chamber;

placing a thermocycling device in contact with the plate to cause the sequence-tagged sites in the chambers to react with genetic material of the organism that is placed in contact with the chambers; and

connecting a computer to the detectors and to the thermocycling device to control operation of the thermocycling device, and to receive signals which correspond to the genetic material of the organism and the sequence-tagged sites of each chamber from the detectors.

12. A method of determining the size of nucleotide sequences of an STR marker contained on genetic material comprising the steps of:

amplifying the nucleotide sequences of the genetic material in a region relating to the STR marker;

performing nucleic acid hybridizations on the amplified nucleotide sequences;

producing signals corresponding to the hybridizations of the amplified nucleotide sequences; and

determining the sizes of the nucleotide sequences contained in the genetic material.

13. A method as described in Claim 12 wherein the hybridizations include a nucleic acid synthesis step.

14. A method as described in Claim 12 wherein the hybridizations include a nucleic acid ligation step.

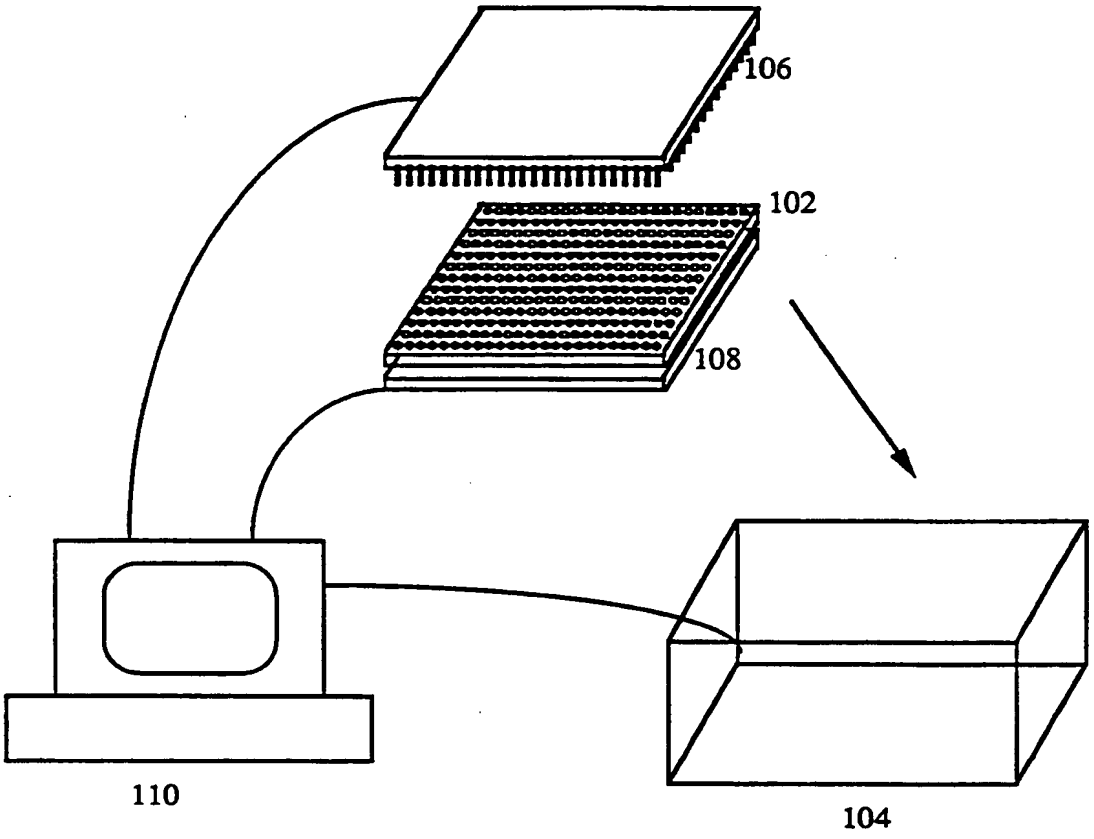


Figure 1.

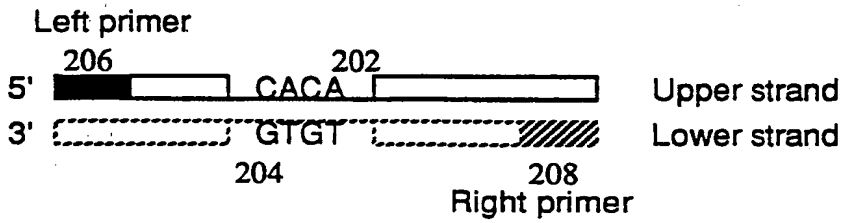

Left primer
206
202
5'
CACA
Upper strand
3'
GTGT
Lower strand
204
208
Right primer

Figure 2.

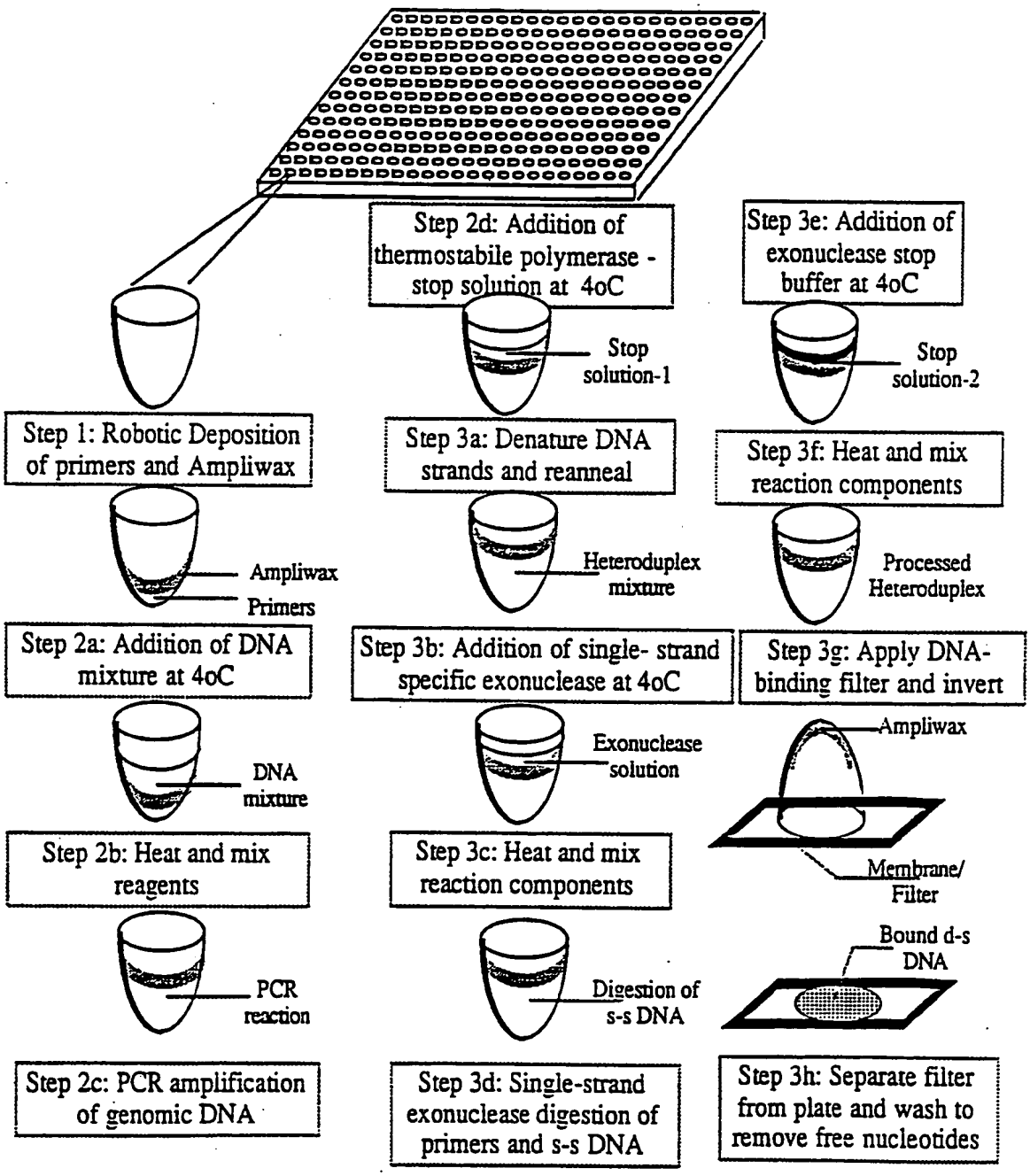

Step 1: Robotic Deposition of primers and Ampliwax
Ampliwax
Primers
Step 2a: Addition of DNA mixture at 4oC
DNA mixture
Step 2b: Heat and mix reagents
PCR reaction
Step 2c: PCR amplification of genomic DNA
Step 2d: Addition of thermostabile polymerase - stop solution at 4oC
Stop solution-1
Step 3a: Denature DNA strands and reanneal
Heteroduplex mixture
Step 3b: Addition of single- strand specific exonuclease at 4oC
Exonuclease solution
Step 3c: Heat and mix reaction components
Digestion of s-s DNA
Step 3d: Single-strand exonuclease digestion of primers and s-s DNA
Step 3e: Addition of exonuclease stop buffer at 4oC
Stop solution-2
Step 3f: Heat and mix reaction components
Processed Heteroduplex
Step 3g: Apply DNA-binding filter and invert
Ampliwax
Membrane/ Filter
Bound d-s DNA
Step 3h: Separate filter from plate and wash to remove free nucleotides

FIGURE 3a

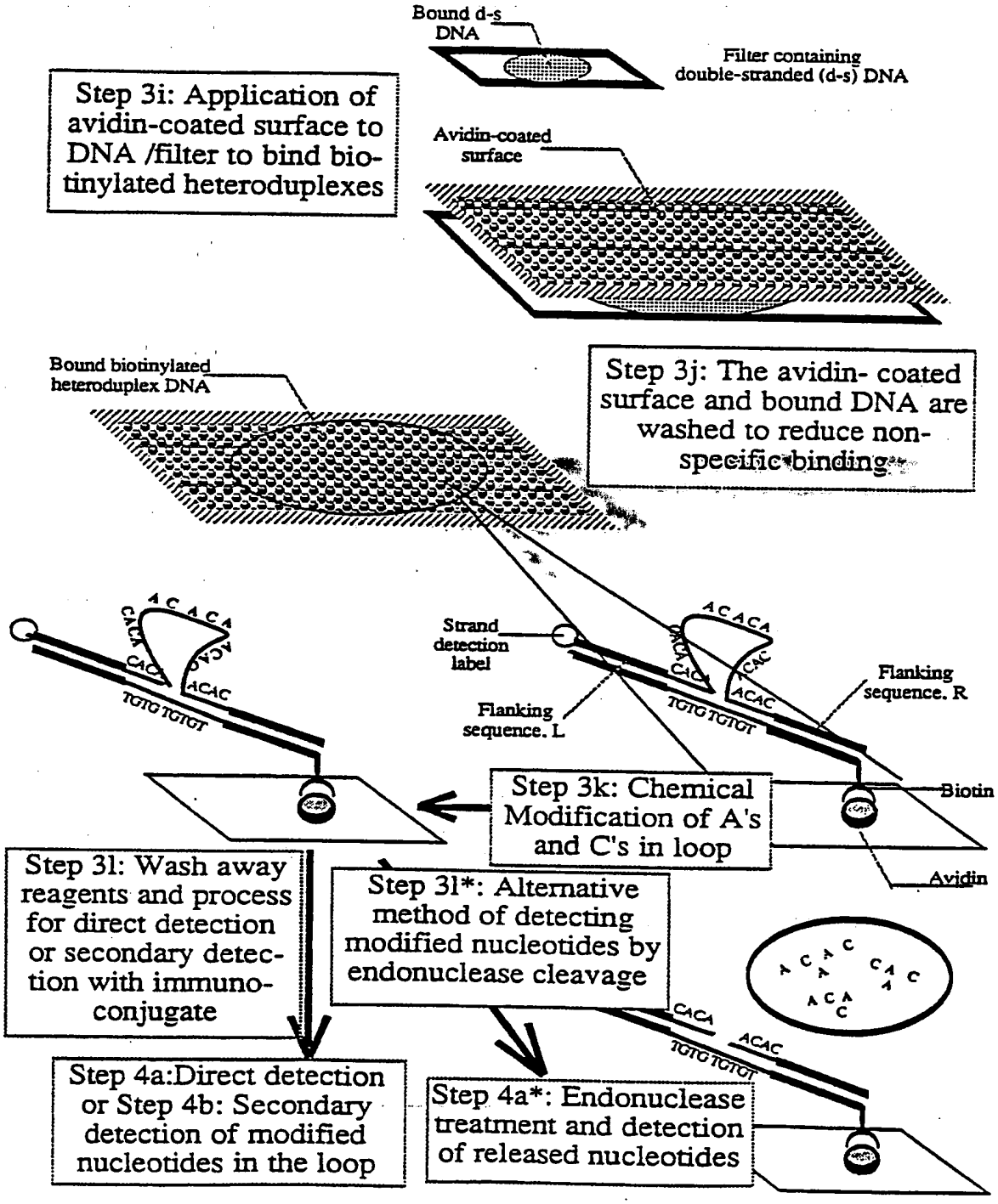
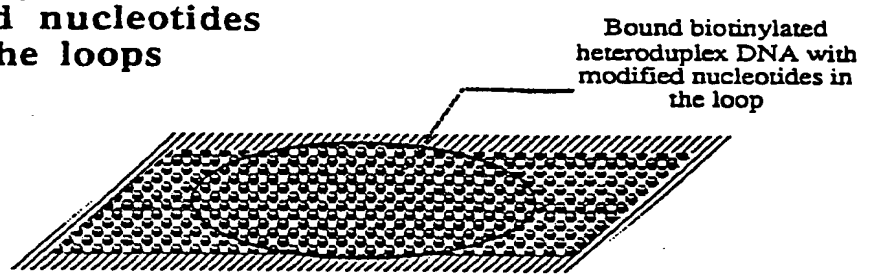
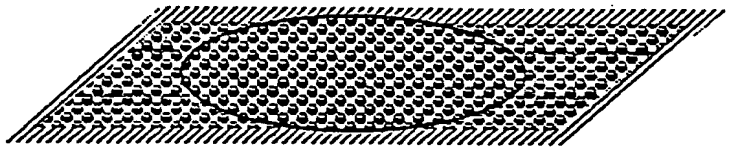

Bound d-s DNA
Filter containing double-stranded (d-s) DNA
Step 3i: Application of avidin-coated surface to DNA /filter to bind biotinylated heteroduplexes
Avidin-coated surface
Bound biotinylated heteroduplex DNA
Step 3j: The avidin- coated surface and bound DNA are washed to reduce non-specific binding
Strand detection label
Flanking sequence. L
Flanking sequence. R
Biotin
Avidin
Step 3k: Chemical Modification of A's and C's in loop
Step 3l: Wash away reagents and process for direct detection or secondary detection with immuno-conjugate
Step 3l*: Alternative method of detecting modified nucleotides by endonuclease cleavage
Step 4a:Direct detection or Step 4b: Secondary detection of modified nucleotides in the loop
Step 4a*: Endonuclease treatment and detection of released nucleotides

FIGURE 3b

Secondary Detection of Modified nucleotides within the loops

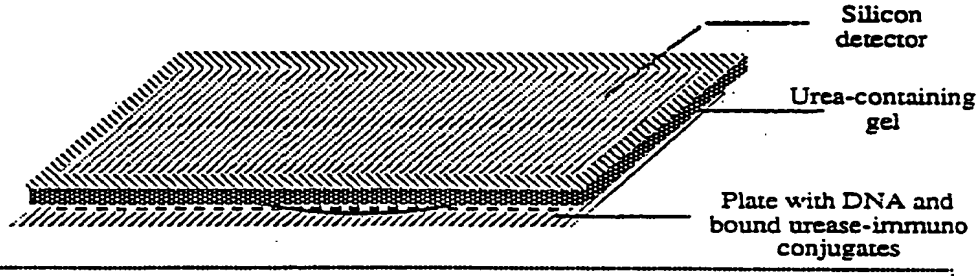


Step 4b: DNA/surface is treated with urease- antibody conjugates that recognize the modified nucleotides (ie: etheno-cytosine, etheno-adenine, OsO4 or carbodiimide derivatives)



Step 4c: Surface is washed to remove unbound antibodies or antisera

Step 4d: A silicon-detector with an overlying thin gel containing urea is placed over the surface



Step 4e: Detection and quantitation of the modified loop nucleotides by the cleavage of urea by the bound urease using the silicon detector to measure the pH shift from the reaction

Step 5: Data Acquisition, Determination of alleles, Data Storage and Analyses

FIGURE 3C

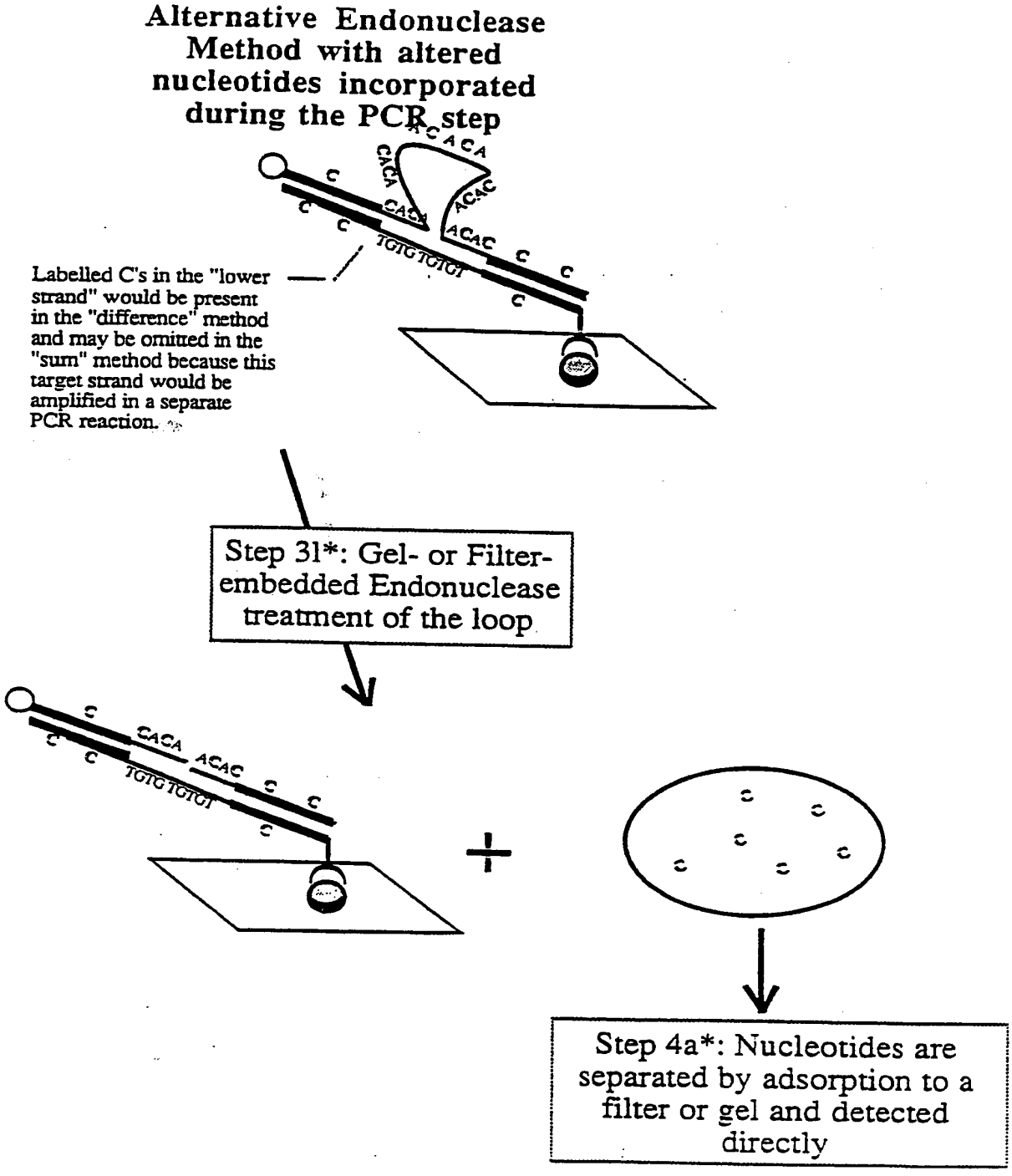

Alternative Endonuclease Method with altered nucleotides incorporated during the PCR step
Labelled C's in the "lower strand" would be present in the "difference" method and may be omitted in the "sum" method because this target strand would be amplified in a separate PCR reaction.
Step 31*: Gel- or Filter-embedded Endonuclease treatment of the loop
Step 4a*: Nucleotides are separated by adsorption to a filter or gel and detected directly

FIGURE 3d

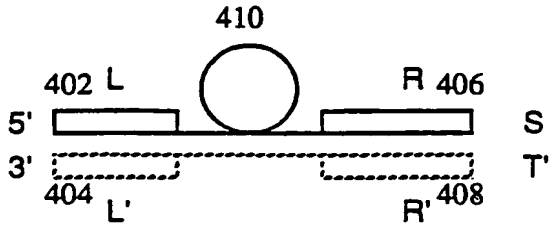


Fig. 4A

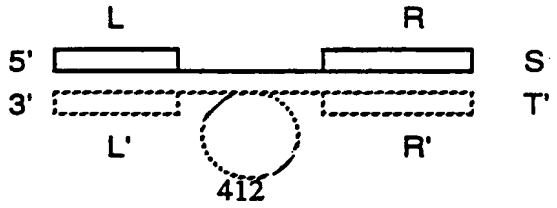


Fig. 4B

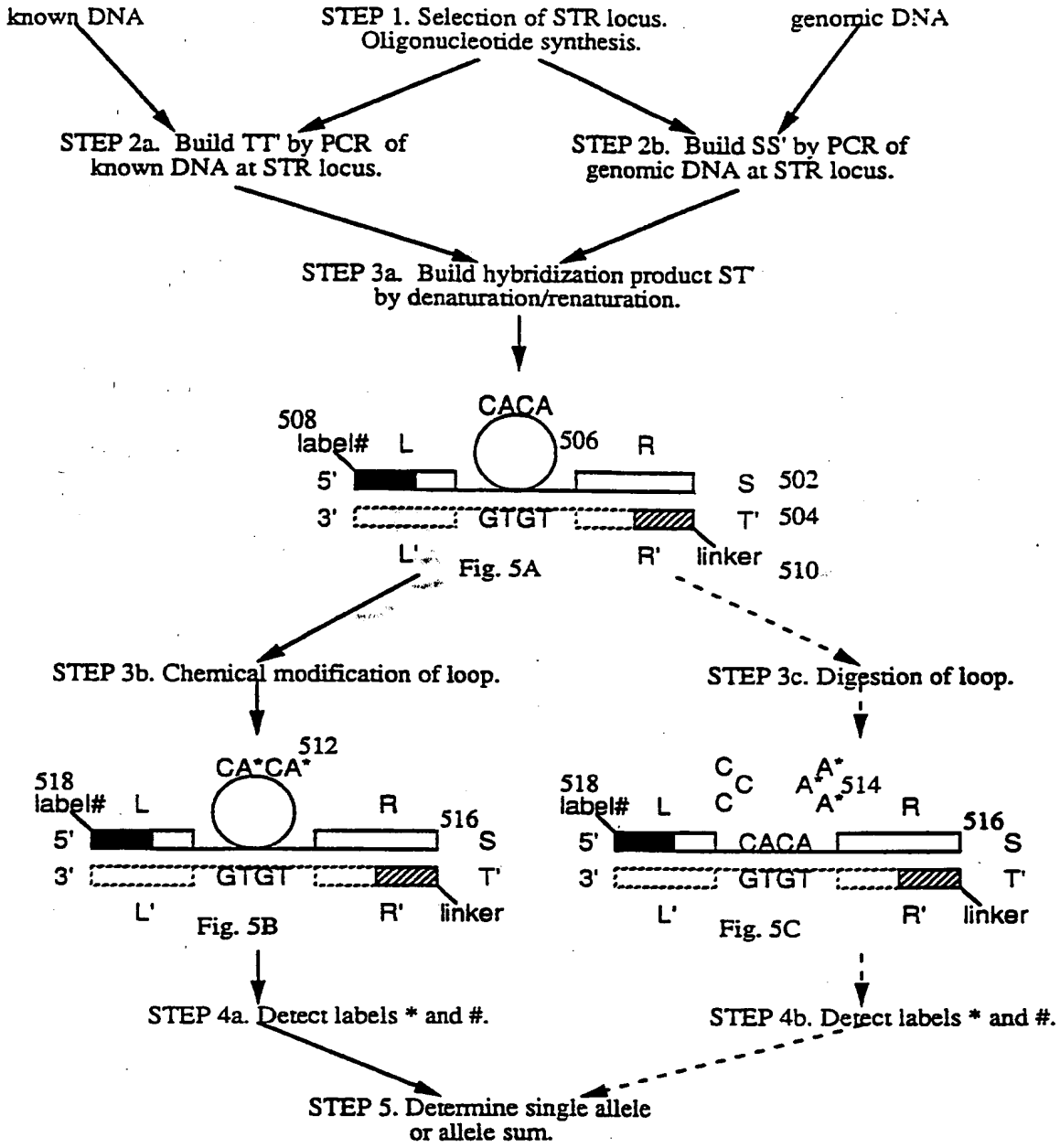

known DNA
STEP 1. Selection of STR locus.
Oligonucleotide synthesis.
genomic DNA
STEP 2a. Build TT' by PCR of known DNA at STR locus.
STEP 2b. Build SS' by PCR of genomic DNA at STR locus.
STEP 3a. Build hybridization product ST' by denaturation/renaturation.
CACA
508
label#
L
506
R
5'
S
502
3'
GTGT
T'
504
L'
R'
linker
510
Fig. 5A
STEP 3b. Chemical modification of loop.
STEP 3c. Digestion of loop.
CA*CA*
512
518
label#
L
R
516
5'
S
3'
GTGT
T'
L'
R'
linker
Fig. 5B
C C A* A* 514
518
label#
L
C
A*
R
516
5'
CACA
S
3'
GTGT
T'
L'
R'
linker
Fig. 5C
STEP 4a. Detect labels * and #.
STEP 4b. Detect labels * and #.
STEP 5. Determine single allele or allele sum.

Figure 5.

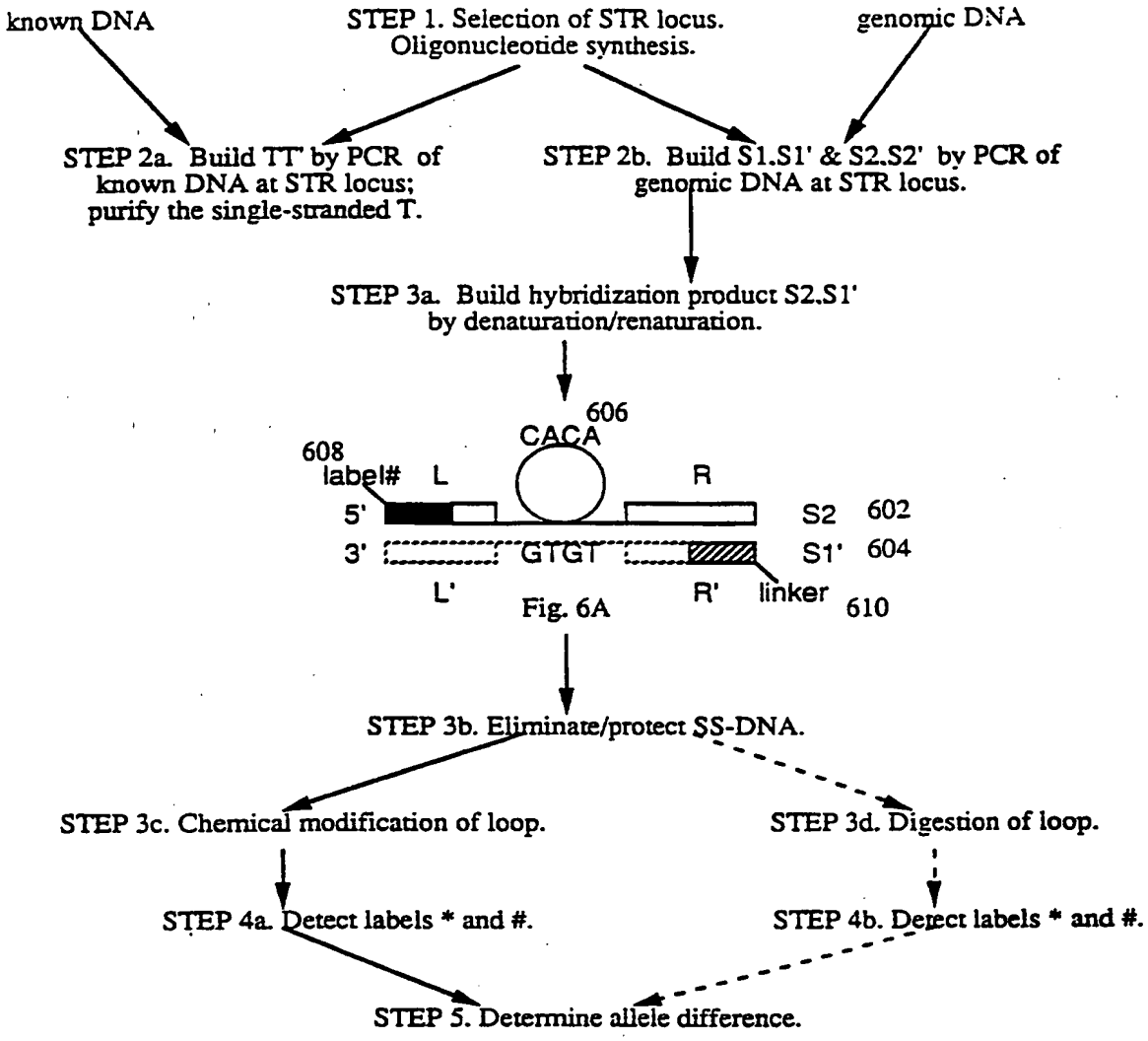

known DNA
STEP 1. Selection of STR locus.
Oligonucleotide synthesis.
genomic DNA
STEP 2a. Build TT' by PCR of known DNA at STR locus; purify the single-stranded T.
STEP 2b. Build S1.S1' & S2.S2' by PCR of genomic DNA at STR locus.
STEP 3a. Build hybridization product S2.S1' by denaturation/renaturation.
606
CACA
608
label#
L
R
5'
S2
602
3'
GTGT
S1'
604
L'
R'
linker
610
Fig. 6A
STEP 3b. Eliminate/protect SS-DNA.
STEP 3c. Chemical modification of loop.
STEP 3d. Digestion of loop.
STEP 4a. Detect labels * and #.
STEP 4b. Detect labels * and #.
STEP 5. Determine allele difference.

Figure 6.

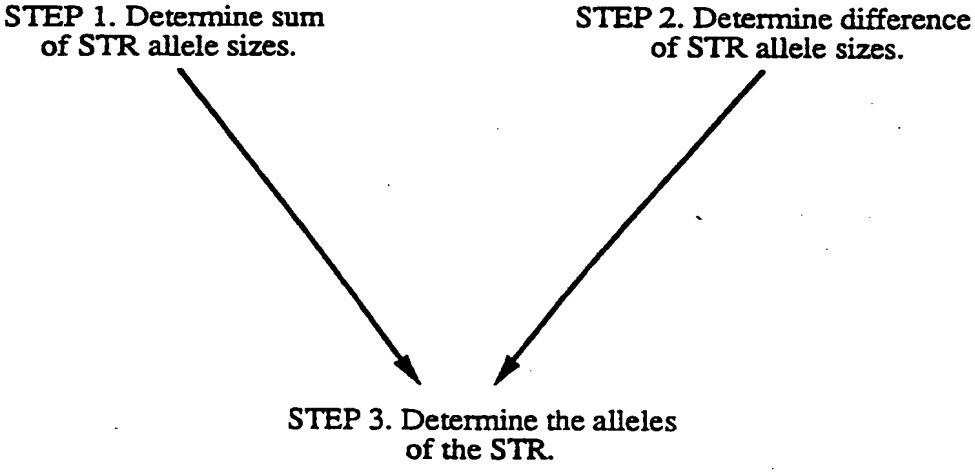

STEP 1. Determine sum of STR allele sizes.
STEP 2. Determine difference of STR allele sizes.
STEP 3. Determine the alleles of the STR.

Figure 7.

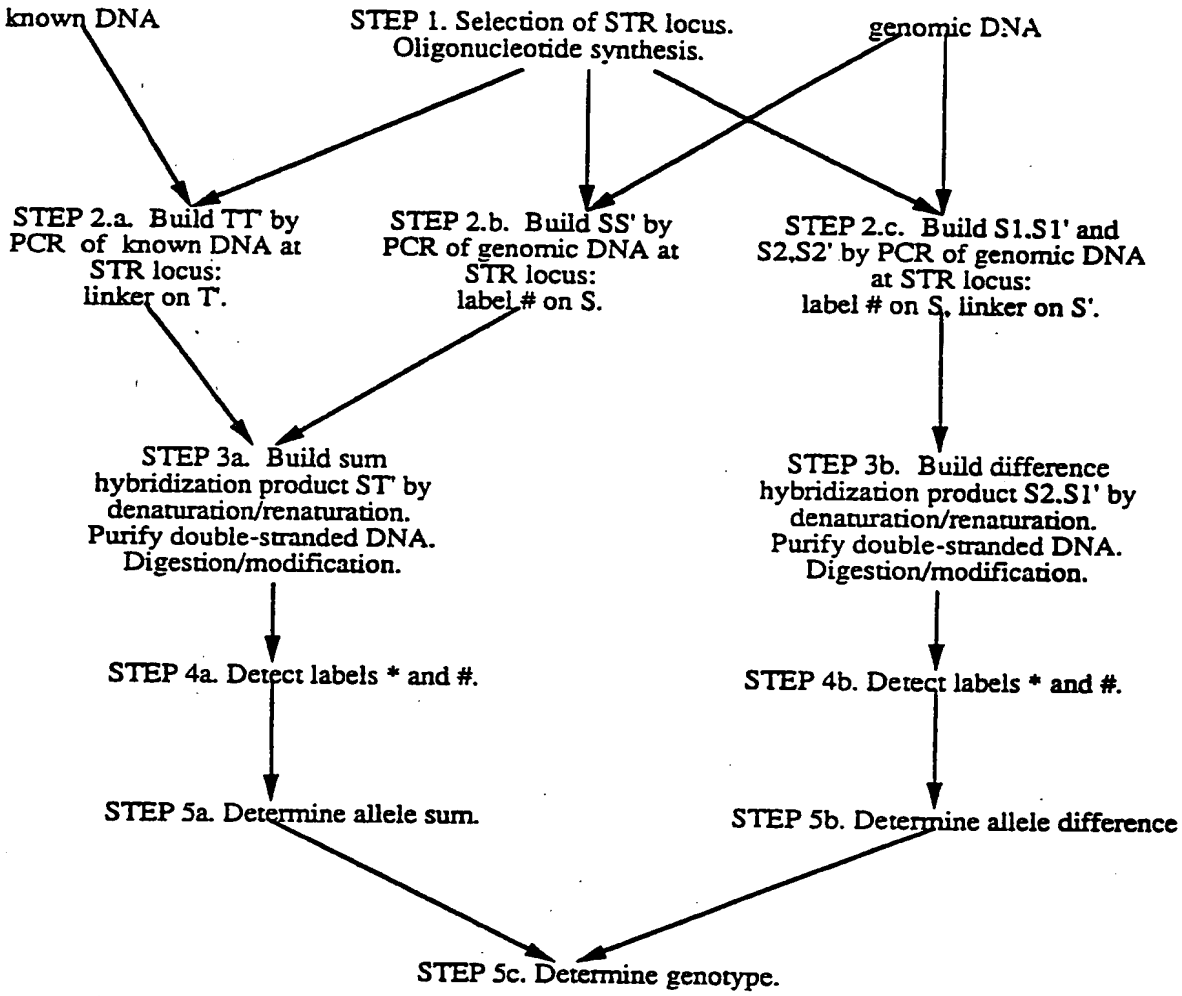

known DNA
STEP 1. Selection of STR locus.
Oligonucleotide synthesis.
genomic DNA
STEP 2.a. Build TT' by PCR of known DNA at STR locus: linker on T'.
STEP 2.b. Build SS' by PCR of genomic DNA at STR locus: label # on S.
STEP 2.c. Build S1.S1' and S2.S2' by PCR of genomic DNA at STR locus: label # on S, linker on S'.
STEP 3a. Build sum hybridization product ST' by denaturation/renaturation. Purify double-stranded DNA. Digestion/modification.
STEP 3b. Build difference hybridization product S2.S1' by denaturation/renaturation. Purify double-stranded DNA. Digestion/modification.
STEP 4a. Detect labels * and #.
STEP 4b. Detect labels * and #.
STEP 5a. Determine allele sum.
STEP 5b. Determine allele difference
STEP 5c. Determine genotype.

Figure 8.

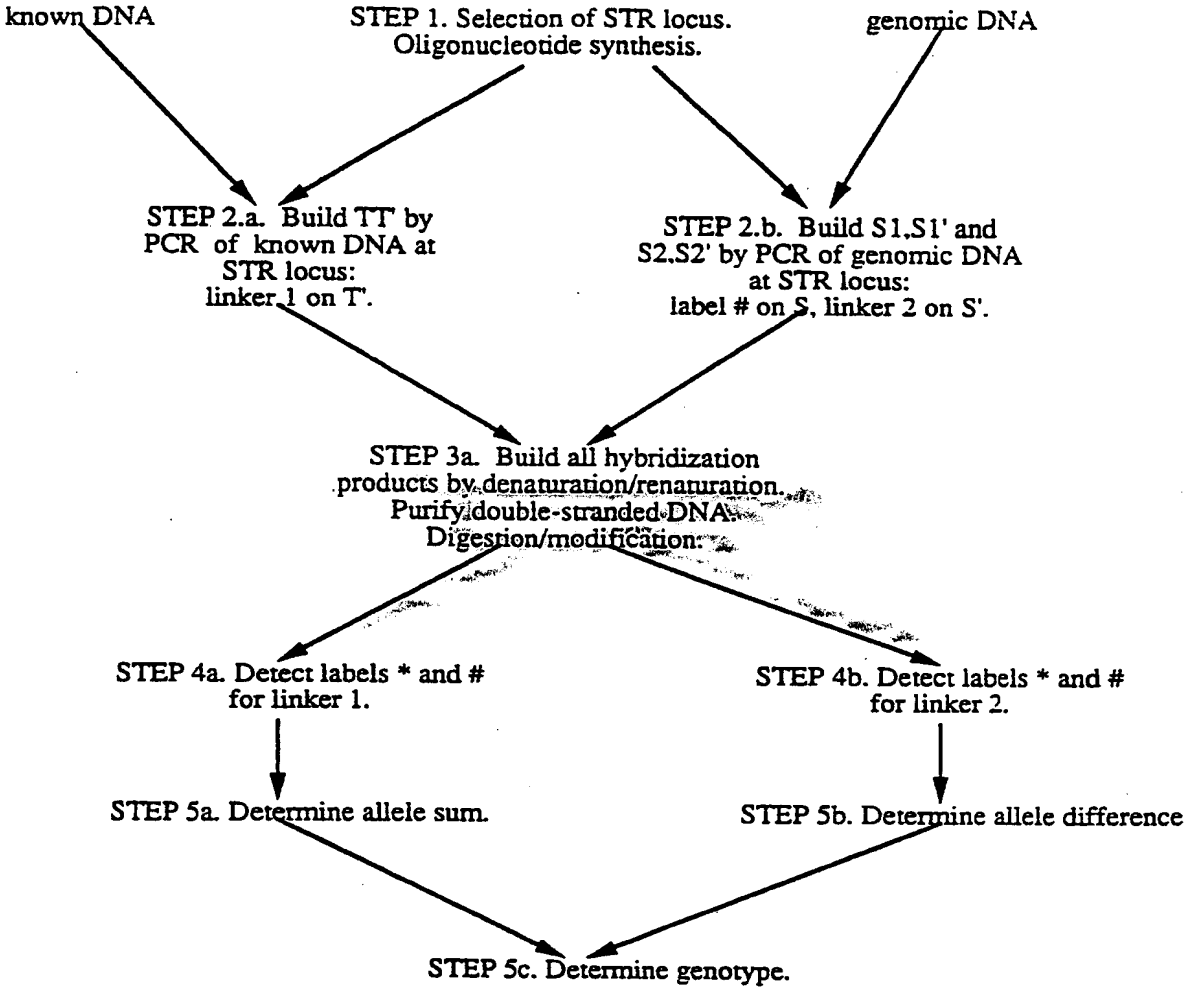

known DNA
STEP 1. Selection of STR locus.
Oligonucleotide synthesis.
genomic DNA
STEP 2.a. Build TT' by PCR of known DNA at STR locus: linker 1 on T'.
STEP 2.b. Build S1,S1' and S2,S2' by PCR of genomic DNA at STR locus: label # on S, linker 2 on S'.
STEP 3a. Build all hybridization products by denaturation/renaturation. Purify double-stranded DNA. Digestion/modification.
STEP 4a. Detect labels * and # for linker 1.
STEP 4b. Detect labels * and # for linker 2.
STEP 5a. Determine allele sum.
STEP 5b. Determine allele difference
STEP 5c. Determine genotype.

Figure 9.

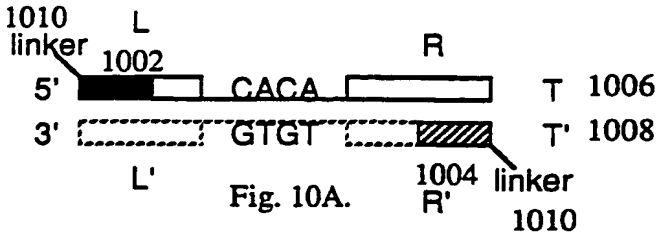


Fig. 10A.

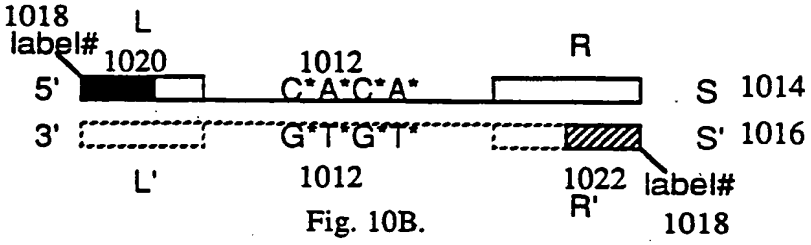


Fig. 10B.

	S' [2n *, 1 #]	T' [linker]
S [2n *, 1 #]	SS': loop=0 2 label # 4n label * 0 linker	ST': loop=s-t 1 label # 2n label * 1 linker
T [linker]	TS': loop=s-t 1 label # 2n label * 1 linker	TT': loop=0 0 label # 0 label * 2 linker

Fig. 10C.

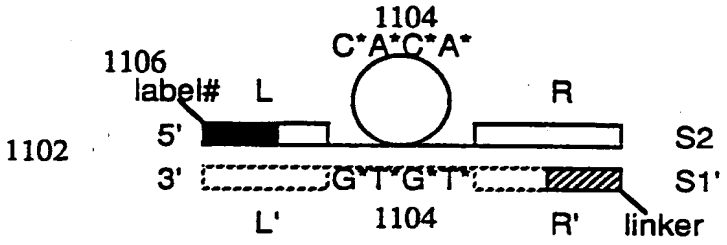


Fig. 11A.

	S1' [2n *, linker]	S2' [2n *, linker]
S1 [2n *, 1 #]	S1S1': loop=0 1 label # 2n*0 label * 1 linker	S1S2': loop=s2-s1 1 label # 2n label * 1 linker
S2 [2n *, 1 #]	S2S1': loop=s2-s1 1 label # 2n label * 1 linker	S2S2': loop=0 1 label # 2n*0 label * 1 linker

Fig. 11B.

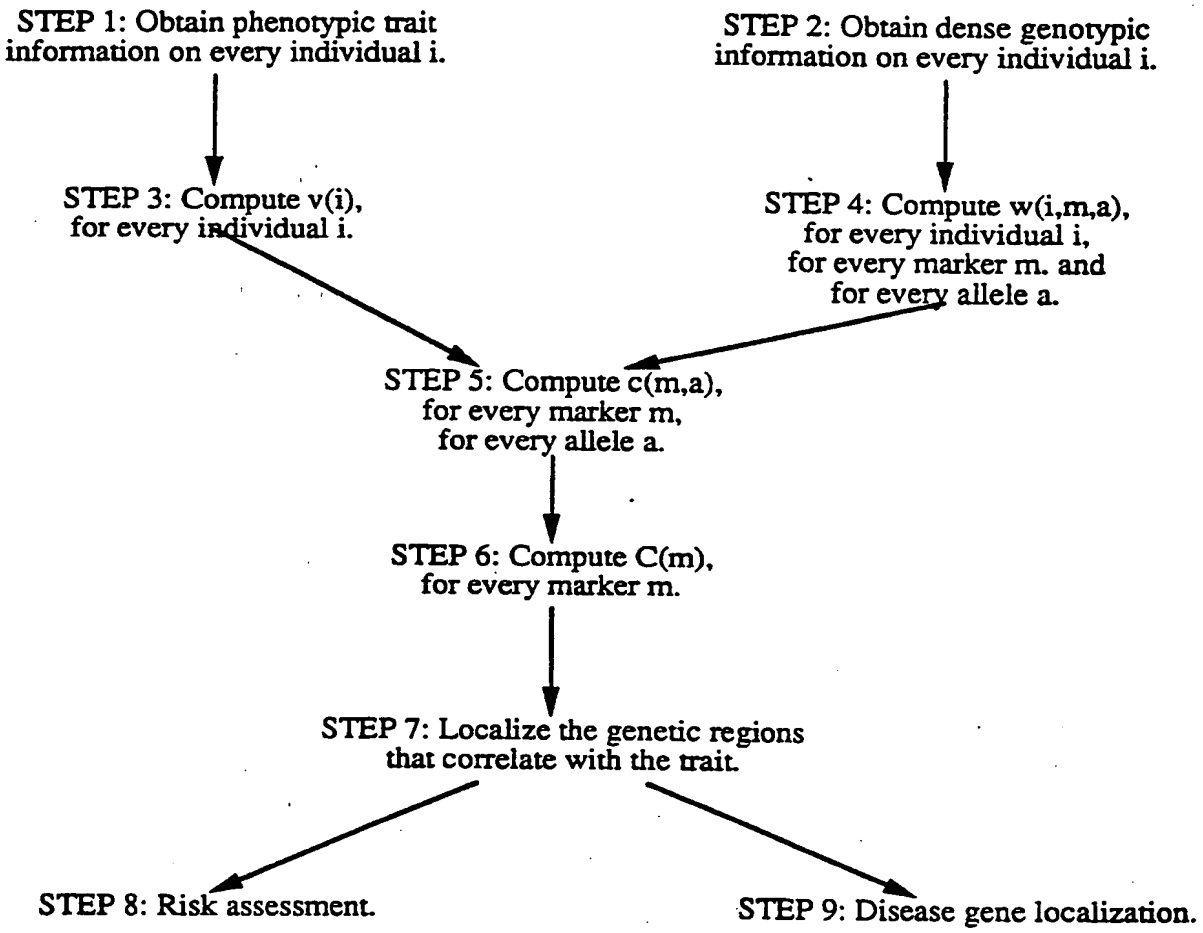

STEP 1: Obtain phenotypic trait information on every individual i.
STEP 2: Obtain dense genotypic information on every individual i.
STEP 3: Compute v(i), for every individual i.
STEP 4: Compute w(i,m,a), for every individual i, for every marker m. and for every allele a.
STEP 5: Compute c(m,a), for every marker m, for every allele a.
STEP 6: Compute C(m), for every marker m.
STEP 7: Localize the genetic regions that correlate with the trait.
STEP 8: Risk assessment.
STEP 9: Disease gene localization.

Figure 12.

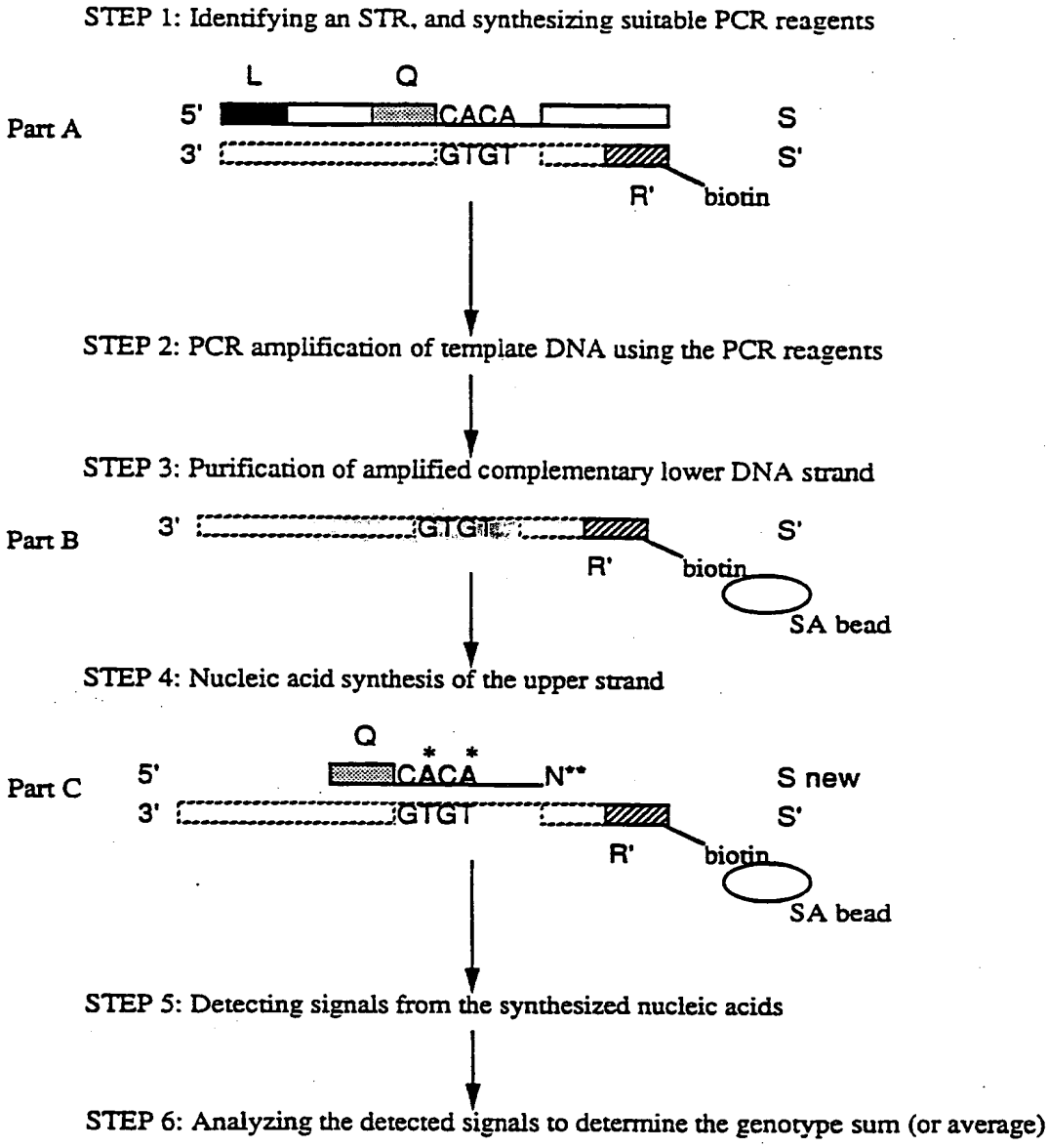

STEP 1: Identifying an STR, and synthesizing suitable PCR reagents
L
Q
Part A
5'
CACA
S
3'
GTGT
S'
R'
biotin
STEP 2: PCR amplification of template DNA using the PCR reagents
STEP 3: Purification of amplified complementary lower DNA strand
Part B
3'
GTGT
S'
R'
biotin
SA bead
STEP 4: Nucleic acid synthesis of the upper strand
Q
Part C
5'
CACA
N**
S new
3'
GTGT
S'
R'
biotin
SA bead
STEP 5: Detecting signals from the synthesized nucleic acids
STEP 6: Analyzing the detected signals to determine the genotype sum (or average)

Figure 13.

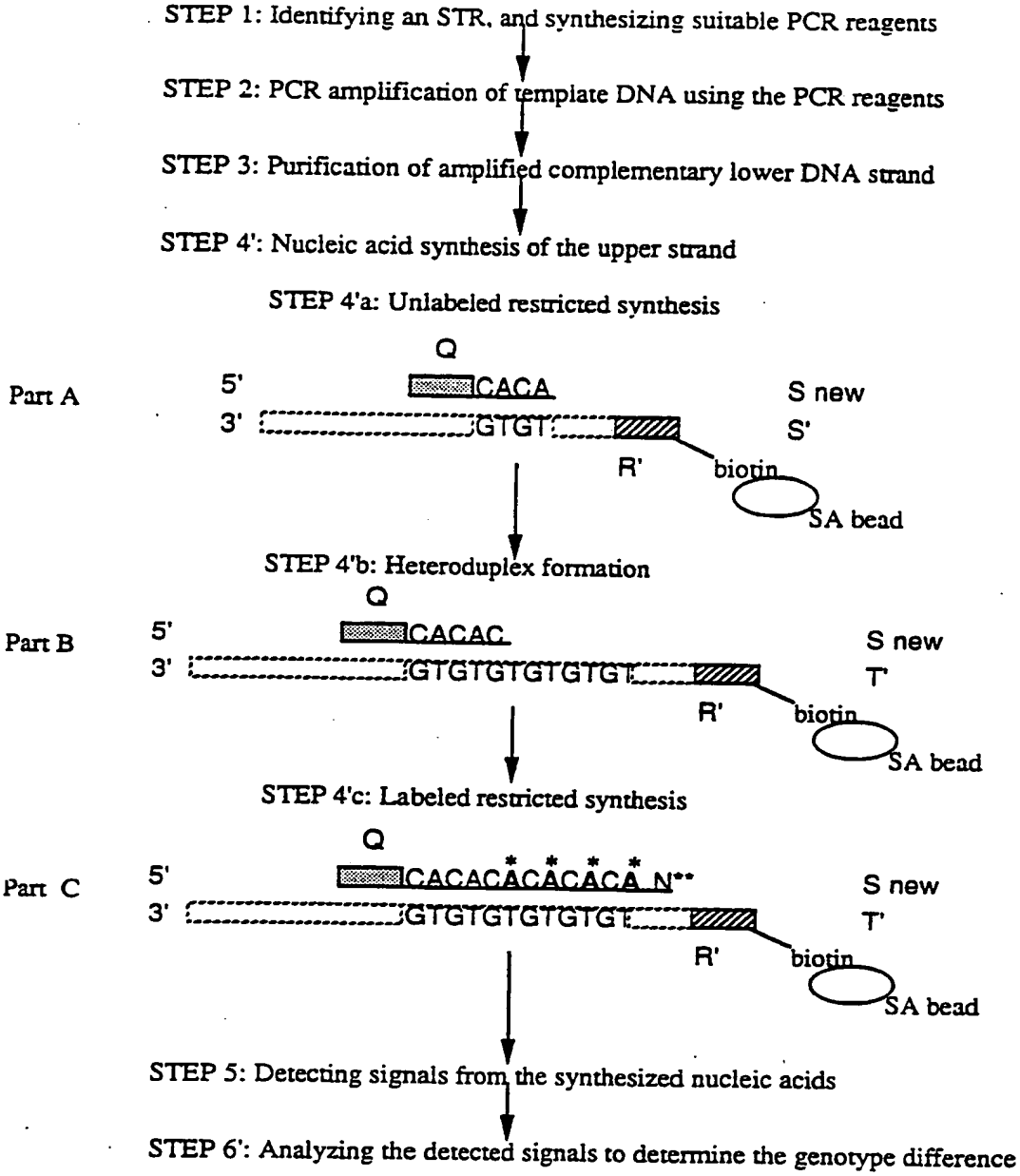

STEP 1: Identifying an STR, and synthesizing suitable PCR reagents
STEP 2: PCR amplification of template DNA using the PCR reagents
STEP 3: Purification of amplified complementary lower DNA strand
STEP 4': Nucleic acid synthesis of the upper strand
STEP 4'a: Unlabeled restricted synthesis
Part A
Q
5'
CACA
S new
3'
GTGT
S'
R'
biotin
SA bead
STEP 4'b: Heteroduplex formation
Part B
Q
5'
CACAC
S new
3'
GTGTGTGTGTG
T'
R'
biotin
SA bead
STEP 4'c: Labeled restricted synthesis
Part C
Q
5'
CACACACACACA N**
S new
3'
GTGTGTGTGTG
T'
R'
biotin
SA bead
STEP 5: Detecting signals from the synthesized nucleic acids
STEP 6': Analyzing the detected signals to determine the genotype difference

Figure 14.

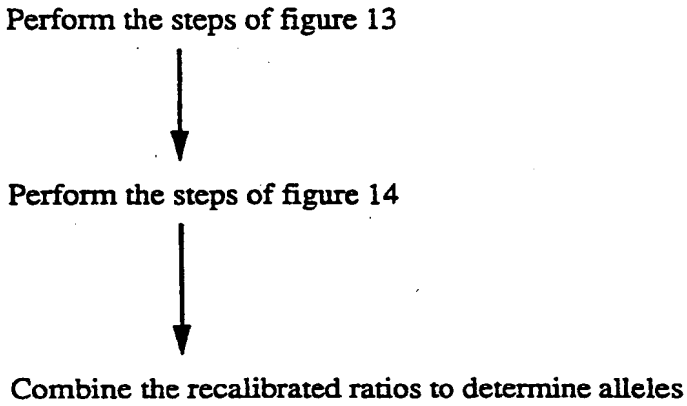

Perform the steps of figure 13
Perform the steps of figure 14
Combine the recalibrated ratios to determine alleles

Figure 15.

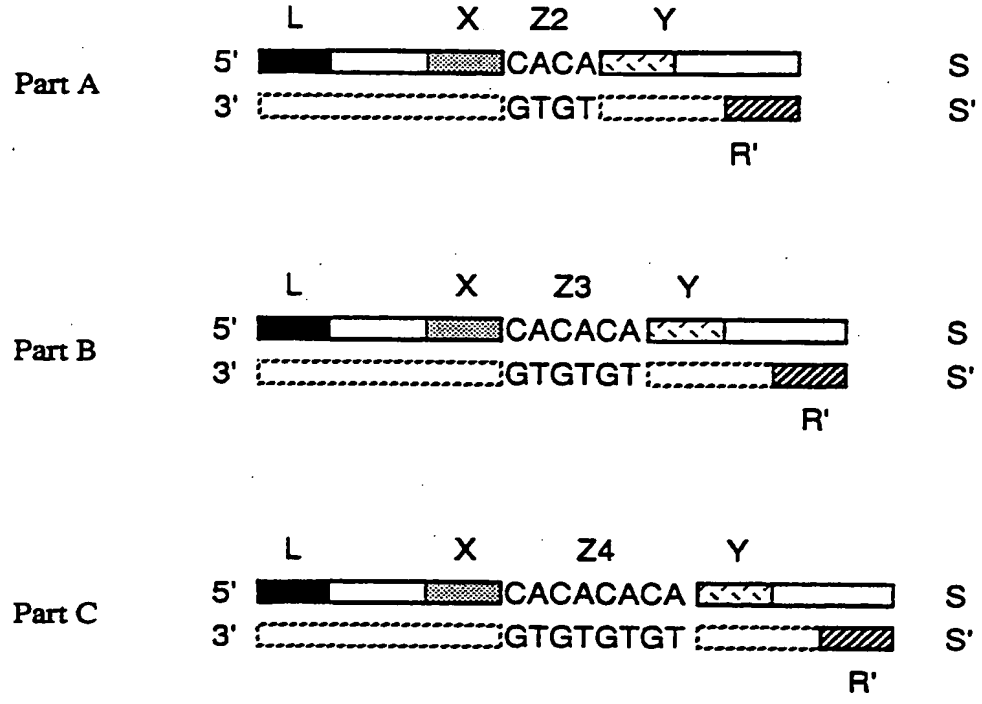
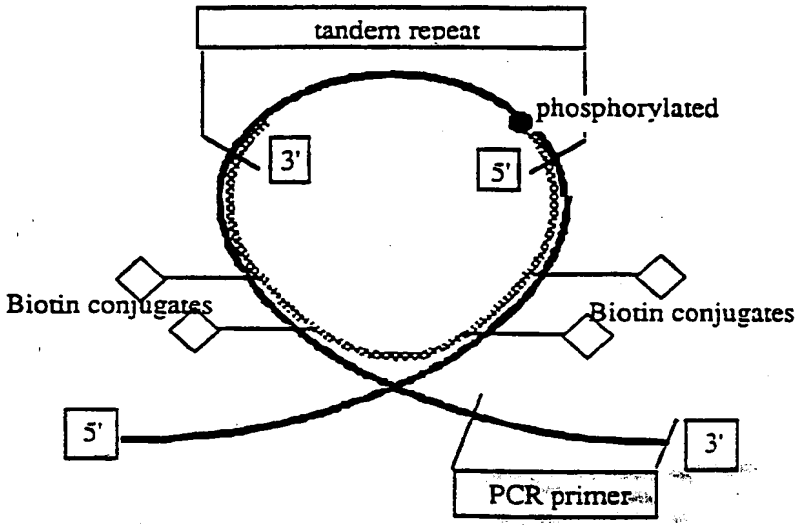

Part A
L
X
Z2
Y
5'
CACA
S
3'
GTGT
S'
R'
Part B
L
X
Z3
Y
5'
CACACA
S
3'
GTGTGT
S'
R'
Part C
L
X
Z4
Y
5'
CACACACA
S
3'
GTGTGTGT
S'
R'

Figure 16.

Part A



Part B

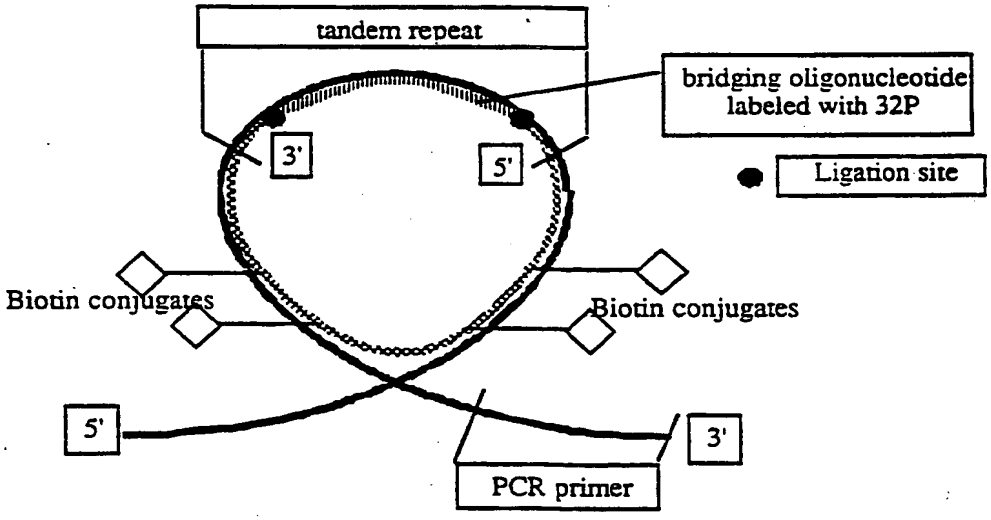


Figure 17.